Structural studies on polyketide enzymes

Patrick Dall'Aglio

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Medical and Veterinary Sciences

Department of Biochemistry

September 2008

Word count: 43135

Dedication

I would like to dedicate this thesis to Giulia and Nerone, who stand by my side every day, and to my parents and grandparents for their precious support.

Abstract

Polyketides are a class of secondary metabolites that includes compounds with antibiotic and anticancer properties. In type I polyketide synthesis, the individual activities are arranged into large polyproteins, while in type II polyketide synthesis enzymes are encoded separately.

Streptomyces coelicolor actinorhodin acyl carrier protein synthase (AcpS) is a promiscuous enzyme involved in type II polyketide synthesis that catalyses the transfer of the phosphopantetheine chain from coenzyme A (CoA) to acyl carrier protein (ACP).

The crystal structures of *S. coelicolor* apo-AcpS and of AcpS in complex with CoA and acetyl-CoA were solved. These holo-AcpSs are the first type I AcpS structures where a magnesium ion, essential for activity, is present in the active site.

Analysis of crystal structures of AcpS in complex with CoA and acetyl-CoA reveals the presence of a defined hydrophobic cavity for the phosphopantetheine chain of CoA. Mobile loop ML2, that defines one side of this cavity, could enlarge this pocket allowing extended CoAs to bind.

A model was made for *S. coelicolor* AcpS:ACP complex and used to interpret fluorescence binding studies of mutant ACPs. Conserved arginines were identified on helix α1 of AcpS which are proposed to interact with conserved aspartate and glutamate residues on helix α2 of ACP. The interaction between AcpS Arg15 and ACP Asp41 could be important for activity and is proposed to lock the ACP serine into a reactive conformation.

His110 and Asp11 in the AcpS active site could deprotonate (in)directly the active ACP serine. Each was mutated to alanine, and the respective mutant proteins were crystallized and characterized biochemically. H110A was inactive but ITC confirmed CoA binding, giving evidence of a possible role in catalysis. D111A had reduced activity. The structure revealed that CoA binds in two alternative conformations, confirmed by ITC measurements, one of which is likely to interfere with catalysis.

Acknowledgements

I would like to thank the following people:

Dr Andrea Hadfield, for her supervising help, suggestions and advice;

Dr Matt Crump and Dr John Crosby, for the kind donation of some acyl carrier protein, acyl carrier protein synthase, actinorhodin and daunorubicin ketoreductase DNA, and for ACP NMR structures;

Dr Russell Cox, for his collaboration and suggestions;

Dr Chris Arthur, for the ACP mutational data and mass-spectroscopy assistance;

Dr Dave Miller, for ITC assistance.

A special thanks also to all members of C56/C58, in particular to May Marsh for sharing ideas and for the help in the lab, Dr Rebecca Conners and Dr Nick Burton for their help in protein crystallography, and to Mike Booth for ideas.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original, except where indicated by special reference in the text, and no part of the dissertation has been submitted for any other academic award. Any views expressed in the dissertation are those of the author.

Signed: Rutik Dalvi Date: 12/01/2009

Abbreviations

6-DEB	6-deoxyerythronolide B
6-MSA	6-methylsalicylic acid
ACP	acyl carrier protein
AcpS	acyl carrier protein synthase
act	actinorhodin
ADP	adenosine diphosphate
AT	acyl transferase
bp	base pair
CCD	charge-coupled device
cDNA	complementary DNA
CIF	chain initiation factor
CLF	chain length factor
CoA	coenzyme A
Da	Dalton
DEBS	6-deoxyerythronolide B synthase
DH	dehydratase
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DpsE	daunorubicin ketoreductase
dsDNA	double-stranded DNA
EDTA	ethylenediaminetetraacetic acid
ER	enoyl reductase
ESMS	electrospray mass spectrometry
FAS	fatty acid synthesis
HPLC	high pressure liquid chromatography
hr	hour
IPTG	isopropyl-β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
kbp	kilobases
kDa	kiloDalton
KR	ketoreductase
KS	ketosynthase
LB	Luria-Bertani
LNKS	lovastatin nonaketide synthase
MAD	multiple wavelength anomalous diffraction
MAT	malonyl-acetyl transferase
MCS	multiple cloning site
MeT	methyl transferase
min	minute
MIR	multiple isomorphous replacement
MPD	2-methyl-2,4-pentanediol
MR	molecular replacement
MT	malonyl transferase
NAD(H)	nicotinamide adenine dinucleotide (reduced form)
NADP(H)	nicotinamide adenine dinucleotide phosphate (reduced form)
NMR	nuclear magnetic resonance

OD_{600}	optical density at 600 nm
PCR	polymerase chain reaction
PEG	polyethylene glycol
PKS	polyketide synthase
RMSD	root mean square deviation
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SQTKS	squalestatin tetraketide synthase
TE	thioesterase
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
UV	ultraviolet
v/v	volume:volume ratio
w/v	weight:volume ratio

Table of Contents

Dedication	ii
Abstract	iii
Acknowledgements	iv
Author's declaration	v
Abbreviations	vi
Table of contents	viii
List of tables	xiii
List of figures	xiv
INTRODUCTION	1
Chapter 1 Introduction	2
1.1 The modern pharmaceutical industry	2
1.2 Polyketide natural products	3
1.3 The fatty acid synthesis	4
1.3.1 FAS biosynthetic pathway	4
1.3.2 Structure of type I mammalian FAS	6
1.3.3 Fungal type I FAS	8
1.3.4 *E. coli* type II FAS	9
1.4 Polyketide biosynthesis	10
1.4.1 Polyketide synthases	12
1.4.2 Type I PKSs	12
1.4.2.1 Iterative type I PKSs	12
1.4.2.1.1 6-Methylsalicylic acid synthase	13
1.4.2.1.2 Lovastatin polyketide synthesis	13
1.4.2.1.3 SQTKS	15
1.4.2.2 Modular type I PKSs	16
1.4.2.2.1 The erythromycin polyketide synthase	16
1.4.3 Type II PKSs	18
1.4.3.1 Type II PKS gene cluster organization	18
1.4.3.2 Acyl Carrier Protein	19
1.4.3.3 Acyl carrier protein Synthase	22
1.4.3.3.1 *Streptomyces coelicolor* AcpS	26
1.4.3.4 Ketosynthase α and ketosynthase β	27
1.4.3.5 Ketoreductases	29
1.4.3.5.1 Actinorhodin ketoreductase	29
1.4.3.6 Cyclases and aromatases	33
1.4.3.7 Tailoring enzymes	34
1.4.3.8 Daunorubicin/doxorubicin synthesis	34
1.4.4 Type III PKSs	36
1.5 Objectives and aims	38

MATERIALS AND METHODS 39

Chapter 2 Methods 40
2.1 Expression vectors 40
2.1.1 pET vectors 40
2.2 *E. coli* BL21 star (DE3) 41
2.3 PCR 42
2.4 Stratagene site-directed mutagenesis 42
2.5 Chromatography techniques 43
2.5.1 Affinity Chromatography 45
2.5.2 Ion exchange chromatography 47
2.5.3 Size exclusion chromatography 47
2.6 Electrophoresis 48
2.7 ITC 49
2.8 Protein crystallography 51
2.8.1 Protein crystallization 51
2.8.2 Vapour diffusion technique 53
2.8.3 X-rays diffraction 55
2.8.3.1 The choice of X-rays 55
2.8.3.2 X-ray diffraction 56
2.8.3.3 Data collection, processing and reduction 58
2.8.3.4 The Fourier transform 59
2.8.3.5 The phase problem 61
2.8.3.6 Molecular replacement 62
2.8.3.7 Refinement and model validation 63
2.8.3.8 Density modification 64

Chapter 3 Materials and protocols 66
3.1 Media 66
3.1.1 Luria-Bertani (LB) liquid medium 66
3.1.2 LB plates 66
3.2 Vectors 66
3.3 Host strain 66
3.4 Molecular biology 67
3.4.1 Agarose gel electrophoresis 67
3.4.2 DNA miniprep 67
3.4.3 Restriction enzyme digest 67
3.4.4 Site-directed mutagenesis for AcpS mutants 67
3.4.5 Heat-shock transformation of plasmid DNA into competen *E. coli* 68
3.4.6 Glycerol stock 69
3.5 Protein expression 69
3.6 Protein purification 69
3.6.1 Buffers 69
3.6.2 Cell harvest and bacterial lysis 69
3.6.3 Affinity chromatography 70
3.6.4 Ion exchange chromatography 71
3.6.5 Size exclusion chromatography (gel filtration) 71
3.6.6 Thrombin cleavage of wt his-AcpS and purification of untagged AcpS 72

3.6.7 Dialysis 72
3.6.8 Desalting 73
3.6.9 Activation and purification of holo-ACP 73
3.7 Biochemical characterization 73
3.7.1 SDS-PAGE gels 73
3.7.2 Protein concentration 74
3.7.3 Bradford assay 74
3.7.4 Mass spectroscopy 75
3.8 X-ray crystallography 75
3.8.1 Crystallization trials 75
3.8.2 Cryoprotectant and loop mounting 75
3.8.3 Data collection and processing 76
3.8.4 Molecular replacement 76
3.8.5 3-fold averaging and density modification for apo-AcpS 76
3.8.6 Refinement and building 77
3.8.7 Generation of AcpS-ACP complex model 77
3.9 ITC 78

RESULTS AND DISCUSSION 79

Chapter 4 The Acyl carrier protein Synthase (AcpS) 80
4.1 Sequence analysis 80
4.2 Expression of wild type AcpS 81
4.3 Purification of wild type AcpS 81
4.4 AcpS crystallization, data collection, processing and refinement 83
4.4.1 Apo-AcpS data collection and processing 83
4.4.2 Molecular replacement and refinement 84
4.4.3 Geometry and B-factor analysis 86
4.5 AcpS in complex with CoA 88
4.5.1 Holo-AcpS crystallization and data collection 88
4.5.2 Molecular replacement and refinement 90
4.5.3 Geometry and B-factor analysis 91
4.6 Structural analysis of the Acyl carrier protein Synthase 92
4.6.1 AcpS fold and oligomerization 92
4.6.2 Active site 93
4.6.2.1 CoA binding site 93
4.6.2.2 Magnesium binding site 94
4.6.3 Comparison between apo- and holo-AcpS 95
4.7 Multiple sequence alignment of analogue type I phosphopantetheinyl transferases 97
4.8 Comparison between *B. subtilis* and *S coelicolor* holo-AcpSs 98
4.9 Comparison between *S. pneumonias* AcpS in complex with 3',5'-ADP and *S. coelicolor* holo-AcpS 100
4.10 Comparison between *S. coelicolor* and human holo-AcpS 102

Chapter 5 Investigating the cofactor promiscuity of AcpS 105
5.1 AcpS in complex with acetyl-, propionyl- and malonyl-CoA 105
5.2 Determination of the structure of the AcpS in complex with acetyl-CoA 106
5.2.1 Acetyl-CoA AcpS crystallization 106

5.2.2 Data collection and processing 107
5.2.3 Molecular replacement and refinement 108
5.2.4 Geometry and B-factor analysis 109
5.3 Analysis of the crystal structure of AcpS in complex with acetyl-CoA 111
5.4 Crystallization of Acps in complex with malonyl- and propionyl-CoA 114
5.5 Mass spectroscopy analysis of cofactors 115
5.5.1 Instability of cofactors 117
5.6 Determination of K_m for CoA and acetyl-CoA by ITC 118
5.7 Cofactor promiscuity of *S. coelicolor* AcpS 120
5.7.1 The hydrophobic cavity surrounding the phosphopantetheine chain 120
5.7.2 Role of mobile loop ML2 121
5.7.3 Modelling modified CoAs into the cavity 122

Chapter 6 Insight into AcpS mechanism and its interaction with ACP 124
6.1 Interaction between AcpS and ACP 124
6.1.1 Proposed ACP recognition and interaction 124
6.1.2 Identification of a common structural motif for ACP binding 126
6.1.3 Analysis of mutational ACP data 128
6.1.3.1 ACP E47A and E53A mutants 128
6.1.3.2 ACP D41A and D41E mutants 129
6.2 Proposed reaction mechanism 129
6.2.1 Initiation of ACP binding 129
6.2.2 *S. coelicolor* AcpS:ACP interaction and serine positioning 131
6.2.3 FAS *B. subtilis* AcpS proposed mechanism 134
6.2.4 The role of Asp111 and His110 in serine deprotonation 136
6.2.4.1 Asp111 137
6.2.4.2 His110 138
6.2.5 Conclusions 139
6.2.5.1 Interaction with ACP 139
6.2.5.2 Proposed reaction mechanism 139
6.3 Purification trials of the AcpS-ACP complex 141
6.3.1 ACP expression 141
6.3.2 ACP purification 141
6.3.3 Activation of holo-ACP 143
6.3.4 Purification trials of the AcpS-ACP complex 145
6.3.5 Crystallization trials 146
6.3.6 Thrombin cleavage of AcpS 147
6.3.7 Conclusions 148

Chapter 7 Structure and activity of AcpS mutants 149
7.1 Generation of AcpS mutants 149
7.2 Primers design and cloning 150
7.3 Expression and purification of H110A and D111A mutants 151
7.4 Investigation of AcpS mutant activities on ACP by mass spectroscopy 153
7.5 Determination of AcpS D111A structure 155
7.5.1 AcpS D111A mutant 155
7.5.2 Data collection and processing 156
7.5.3 Molecular replacement and refinement 156
7.5.4 Geometry and B-factor analysis 159

7.6 Determination of AcpS H110A structure 160
7.6.1 AcpS H110AA crystallization 160
7.6.2 Data collection and processing 161
7.6.3 Molecular replacement and refinement 162
7.6.4 Geometry and B-factor analysis 162
7.7 Analysis of the crystal structure of the D111A AcpS mutant 164
7.8 Analysis of the crystal structure of the H110A AcpS mutant 165
7.9 Investigation of CoA binding with ITC 166
7.10 Role of the His110 and Asp111 in catalysis 169
7.10.1 Role of His110 170
7.10.2 Role of Asp111 171
7.10.3 Catalysis revisited 172

Conclusions 173

Appendix 1 Amino Acids 177
Appendix 2 Vector maps 180
pET-11a-d 180
pET-15b 181
Appendix 3 Crystallization kits 182
Crystal Screens 1 and 2 (Hampton Research) 182
PEG/Ion Screen (Hampton Research) 183
Clear Strategy Screen I (MDL) 184
Clear Strategy Screen II (MDL) 184
Structure Screen I and II 96 HT (MDL) 185
PACT-*premiere* 96 HT 187
JCSG-*plus* 96 HT (MDL) 189

Bibliography 191

List of Tables

Chapter 3 Materials and protocols

Table 3.1 Primers used for PCR in AcpS site-directed mutagenesis 68
Table 3.2 Lysis buffers used for each different protein 70
Table 3.3 Equilibration and elution buffers used with the Ni^{2+} affinity column 70
Table 3.4 Ion exchange buffers used with the ion exchange chromatography 71
Table 3.5 Gel filtration buffers used for gel filtration purification 71
Table 3.6 Dialysis buffer used 72
Table 3.7 Desalting buffers used for each different protein 73
Table 3.8 Search models used for the molecular replacement 76

Chapter 4 The Acyl carrier protein Synthase (AcpS)

Table 4.1 Data processing and refinement statistics for the apo-AcpS 86
Table 4.2 Holo-AcpS crystallization conditions 88
Table 4.3 Data processing and refinement statistics for AcpS in complex with CoA 90

Chapter 5 Investigating the cofactor promiscuity of AcpS

Table 5.1 Crystallization conditions for AcpS in complex with acetyl-CoA 106
Table 5.2 Statistic data for AcpS in complex with acetyl-CoA 109
Table 5.3 Data collection statistics for AcpS incubated with propionyl- or malonyl-CoA 115
Table 5.4 Binding affinity for the holo-AcpS and the acetyl-CoA AcpS 118

Chapter 6 Insight into AcpS mechanism and its interaction with ACP

Table 6.1 Affinity for apo-AcpS and percentage of conversion of C17S ACP mutants 128
Table 6.2 Cocrystallization condition for AcpS and holo-ACP 146

Chapter 7 Structure and activity of AcpS mutants

Table 7.1 D111AcpS crystallization conditions 155
Table 7.2 Data processing and refinement statistics for D111A AcpS mutant 158
Table 7.3 Crystallization conditions for the H110A AcpS mutant 161
Table 7.4 Data processing and refinement statistics for D111A AcpS mutant 162
Table 7.5 Affinity constants and stoichiometry for the H110A and D111A mutants 167

List of Figures

Chapter 1 Introduction

Figure 1.1 Structural and functional diversity of polyketides 3
Figure 1.2 Fatty acid biosynthesis 5
Figure 1.3 Model of the functional contacts between FAS domains 7
Figure 1.4 Organisation of the FAS active sites 8
Figure 1.5 The *S. cerevisiae* type I FAS 9
Figure 1.6 Claisen condensation between the starter and the extender units 10
Figure 1.7 Polyketide synthesis 11
Figure 1.8 6-MSA polyketide synthesis 13
Figure 1.9 Lovastatin polyketide synthesis 14
Figure 1.10 Squalestatin S1 and its chain B 15
Figure 1.11 Domain organization of the erythromycin polyketide synthase 17
Figure 1.12 Double helical "Cambridge model" for the erythromycin synthase 18
Figure 1.13 Organisation of gene cluster for type II PKSs 19
Figure 1.14 Phosphopantetheinylation of apo-ACP 20
Figure 1.15 Structures of different apo-ACPs 21
Figure 1.16 Surfactin syntethase-activating enzyme Sfp 23
Figure 1.17 Cartoon representation of the *B. subtilis* FAS AcpS:ACP complex 24
Figure 1.18 Interaction between *B. subtilis* AcpS and ACP 25
Figure 1.19 Crystal structure of the human AcpS in complex with type I fatty acid ACP domain and CoA 26
Figure 1.20 Type II actinorhodin KS_α/ KS_β 27
Figure 1.21 Reaction catalyzed by the KSα and KS_β complex 28
Figure 1.22 Putative biosynthetic scheme for the early steps of actinorhodin biosynthesis 30
Figure 1.23 Actinorhodin ketoreductase crystal structure 31
Figure 1.24 KR open and closed forms 32
Figure 1.25 Daunorubicin/doxorubicin polyketide synthase 35
Figure 1.26 Reactions catalyzedd by type III chalcone and stilbene polyketide synthase 36
Figure 1.27 Proposed mechanism for the chalcone synthase 37

Chapter 2 Methods

Figure 2.1 Expression of a target protein cloned into a pET vector 41
Figure 2.2 The Van Deemter equation 45
Figure 2.3 The structure of the IMAC resin used 46
Figure 2.4 The ITC constant c 50
Figure 2.5 The solubility curve 53
Figure 2.6 The nucleation zone 54
Figure 2.7 The hanging drop and sitting drop technique with vapour diffusion. 55
Figure 2.8 Condition that produces diffracted ray according to the Bragg's law. 57
Figure 2.9 Ewald's sphere of reflection 58

Chapter 4 The Acyl carrier protein Synthase (AcpS)
Figure 4.1 The sequence of the *S. coelicolor* his_6-AcpS 80
Figure 4.2 AcpS expression 81
Figure 4.3 AcpS affiniy chromatography 82
Figure 4.4 AcpS gel filtration chromatography 83
Figure 4.5 Apo-AcpS crystal and related diffraction pattern 84
Figure 4.6 Map improvement after 3-fold averaging 85
Figure 4.7 Ramchandran plot related to the apo-AcpS 87
Figure 4.8 Apo-AcpS B-factor analysis 88
Figure 4.9 Holo-AcpS crystals and example of diffraction image 89
Figure 4.10 Ramchandran plot related to the holo-AcpS 91
Figure 4.11 Holo-AcpS B-factor analysis 92
Figure 4.12 Structure of *S. coelicolor* AcpS 93
Figure 4.13 Electron density map at 1.2 sigma of CoA and the magnesium coordination sphere 94
Figure 4.14 CoA lies at the interface between 2 monomers 96
Figure 4.15 Superimposition between *S. coelicolor* holo- and apo-AcpS 96
Figure 4.16 Multiple sequence alignment of type I AcpSs 97
Figure 4.17 Cα superimposition between *S. coelicolor* and *B. subtilis* holo-AcpS 98
Figure 4.18 Surface representations for CoA binding pockets 99
Figure 4.19 Comparison between ML2 of different proteins 101
Figure 4.20 Superimposition between the human and *S. coelicolor* AcpS active sites 102
Figure 4.21 Comparison between different ML2 equivalent regions 103

Chapter 5 Investigating the cofactor promiscuity of AcpS
Figure 5.1 AcpS hydrophobic cavity 105
Figure 5.2 Cofactor promiscuity of *S. coelicolor* AcpS 106
Figure 5.3 Acetyl-CoA AcpS crystals 107
Figure 5.4 Suggested presence of CoA in the acetyl-CoA AcpS active site 108
Figure 5.5 Ramchandran plot related to the acetyl-CoA AcpS 110
Figure 5.6 Acetyl-CoA AcpS B-factor analysis 111
Figure 5.7 Comparison between the holo-AcpS and acetyl-CoA AcpS active sites 112
Figure 5.8 Conformation of ML2 after CoA and acetyl-CoA binding 113
Figure 5.9 Crystals of AcpS in the presence of malonyl- and propionyl-CoA 114
Figure 5.10 Mass spectra of CoA, acetyl- , propionyl- and malonyl-CoA 116
Figure 5.11 ITC data showing CoA and acetyl-CoA binding to wild type AcpS 119
Figure 5.12 Blanks for CoA and acetyl-CoA binding 119
Figure 5.13 Charged surface representation of the hydrophobic cavity and ML2 surrounding CoA and acetyl-CoA 121
Figure 5.14 Modelling extended CoAs inside the hydrophobic cavity 122

Chapter 6 Insight into AcpS mechanism and its interaction with ACP
Figure 6.1 *B. subtilis* AcpS-ACP interaction 124
Figure 6.2 Multiple sequence alignment of polyketide and fatty acid ACPs 125
Figure 6.3 Multiple sequence alignment of polyketide and fatty acid AcpSs 125
Figure 6.4 Moded of *S. coelicolor* ACP interacting with AcpS 127

Figure 6.5 Proposed ACP binding site for KR 127
Figure 6.6 Initiation of ACP binding 130
Figure 6.7 AcpS:ACP active site 132
Figure 6.8 Positioning of Ser42 in the AcpS active site 133
Figure 6.9 Displacement fo the loop containing the serine 134
Figure 6.10 Reaction mechanism proposed for the *B.subtilis* fatty acid complex 135
Figure 6.11 Possible catalytic role for AcpS Asp111 and His110 137
Figure 6.12 ACP serine deprotonation 138
Figure 6.13 Proposed reaction mechanism 140
Figure 6.14 Amino acid sequence of the actinorhodin C17S ACP 141
Figure 6.15 ACP ion exchange purification 142
Figure 6.16 ACP gel filtration purification 143
Figure 6.17 Mass spectra obtained for the apo-ACP and holo-ACP 144
Figure 6.18 Activation of holo-ACP 145
Figure 6.19 Crystals obtained after co-crystallization of AcpS and holo-ACP 147
Figure 6.20 Thrombin cleavage of AcpS histidine tag 148

Chapter 7 Structure and activity of AcpS mutants
Figure 7.1 Alignments of multiple polyketide synthesis and fatty acid synthesis AcpS sequences 150
Figure 7.2 Alignments between the wild type AcpS and the D111A or H110A mutant 151
Figure 7.3 Purification of AcpS mutants 152
Figure 7.4 Purification of holo-ACP by gel filtration 153
Figure 7.5 Activation of ACP by mutants 154
Figure 7.6 Example of D111A AcpS crystal and its difraction image 155
Figure 7.7 Electron density corresponding to the D111A mutation 156
Figure 7.8 Possible alternative conformations of CoA bound to D111A AcpS 157
Figure 7.9 Ramchandran plot related to D111A mutant 159
Figure 7.10 D111A B-factor analysis 160
Figure 7.11 Example of H110A AcpS crystal and its diffraction image 161
Figure 7.12 Ramchandran plot related to D111A mutant 163
Figure 7.13 H110A B-factor analysis 164
Figure 7.14 D111A AcpS crystal structure 164
Figure 7.15 The H110A mutation in the corresponding crystal structure 166
Figure 7.16 ITC graphs showing the presence of a strong binding site for CoA 168
Figure 7.17 Discovering a second binding site for CoA for the aspartate mutant 169
Figure 7.18 The alternative CoA conformation 171
Figure 7.19 Model for ACP binding to D111A mutant 172

INTRODUCTION

Chapter 1

Introduction

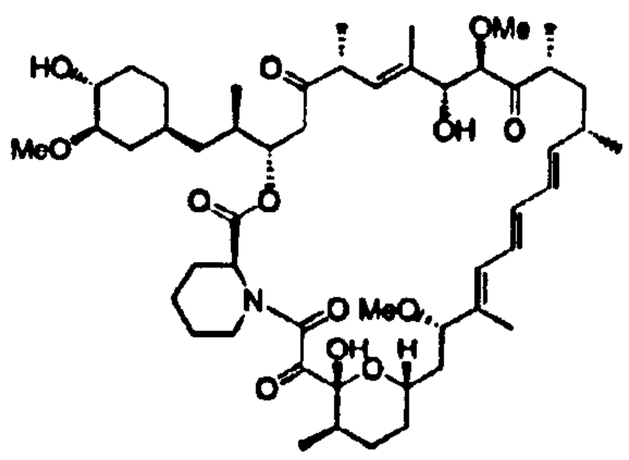
1.1 The modern pharmaceutical industry

Pharmacology as we know it is a relatively recent field of investigation. It is only at the end of the Nineteenth century, with the discovery that an active compound could be not only reproduced in laboratory, but also modified so that it could trigger a specific physiological response, that an incredible leap forward in the treatment of many pathologies was made. Until that moment, treatment of diseases had been based solely on the use of mineral, vegetable and animal extracts. In a very prolific period for the chemical and pharmaceutical fields, the crucial event was the discovery of penicillin by Alexander Fleming in 1928. When he came back from a long holiday, the young Scottish researcher noticed that, on a plate previously inoculated with *Staphilococcus* bacteria, there was a zone around an invading fungus, the *Penicillum notatum*, where bacteria could not grow. He isolated and purified the extract, and investigated the antibacterial properties of the active agent that he called penicillin. Fleming published his discovery in 1929 in the British Journal of Experimental Pathology [1], but little attention was initially paid to his article. A few years later, Fleming's findings were elaborated further by Walter Florey and Ernst Chain. The two scientists managed to improve the purification protocol of penicillin in a stable form, and, after several clinical trials, they developed methods for mass production and distribution of the first antibiotic in 1945. For their massive contribution to the pharmaceutical industry, Fleming, Florey and Chain received the Nobel Prize in Medicine in 1945.

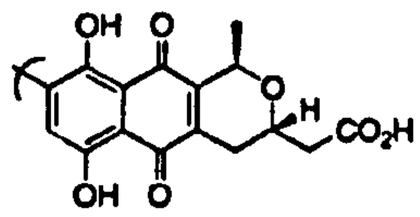
Since the discovery of the first antibiotic, things have progressed quickly and nowadays there is a huge demand for new compounds, and an increasing interest in the possibilities they offer. Moreover, a number of classes of natural products, synthesized in bacteria and fungi [2], with similar properties to those of synthetic compounds, have been identified.

1.2 Polyketide natural products

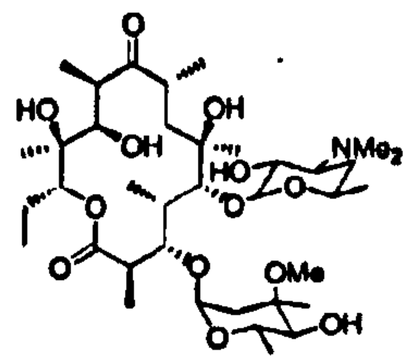
Polyketide products are a broad class of secondary metabolites that exhibit great functional and structural diversities (Figure 1.1), and have a broad range of uses, including antibiotic (erythromycin A, rifamycin S), immunosuppressive (rapamycin), antifungal (amphotericin B), antiparasitic (avermectin), insecticidic (spinosyn A), as well as anticancer (doxorubicin) properties [3]. So, there is a great commercial interest in the production of polyketides because of their wide application for drug discovery.



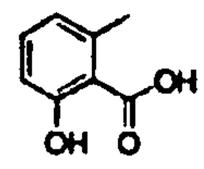
Rapamycin
polyene immunosuppressant



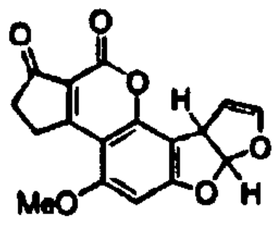
Actinorhodin
aromatic antibiotic



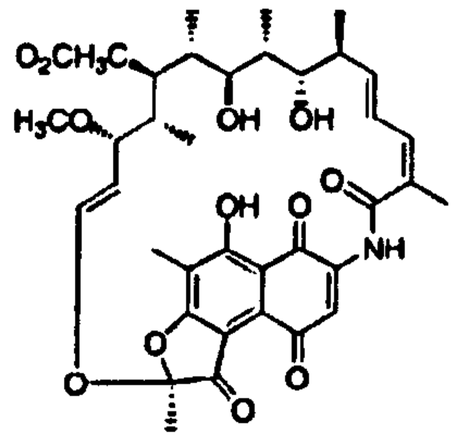
Erythromycin A
macrolide antibiotic



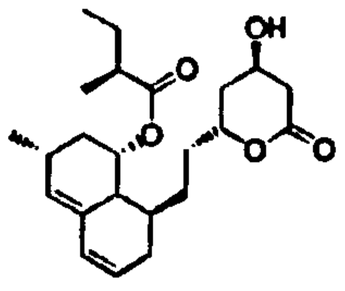
6-Methylsalicylic acid
aromatic antibiotic



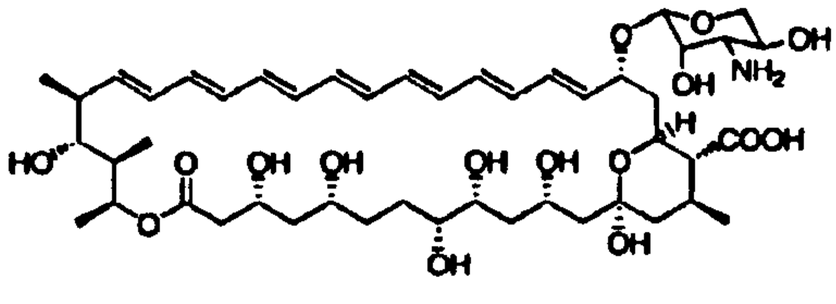
Aflatoxin B1
mycotoxin carcinogen



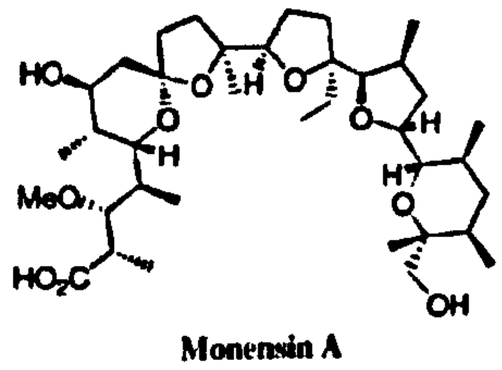
Rifamycin S
ansamycin antibiotic



Lovastatin
anti-cholesterol



Amphotericin B
polyene antifungal



Monensin A
polyether antibiotic

Figure 1.1 Structural and functional diversity of polyketides. (Figure adapted from [3]).

Since the serendipitous discovery of the aromatic orcinol by James Collie in 1893, the polyketide field has expanded considerably. The structures of numerous molecules have been elucidated, and in addition, the cloning, isolation and

characterization of the enzymes responsible for their biosynthesis has begun. Polyketides are synthesized by a broad range of organisms, including bacteria, fungi, marine organisms and higher plants, via the iterative condensation of simple carboxylic acid precursors, a reaction similar to that of fatty acid biosynthesis. The most abundant source of polyketides is the *Actinomycetales*, in particular the *Streptomycetes* and *Saccharopolyspora* strains. *Myxobacteria* and *Pseudomonads* are also other important sources of bacterial polyketides. The precisely programmed polyketide synthases, in association with other chemical modifications that occur after the synthesis of the polyketide chain by the so-called tailoring enzymes, are responsible for the great chemical diversity.

Early experiments revealed that the polyketide synthesis is very similar to fatty acid synthesis (FAS). The two pathways are characterized by a common pool of simple precursors and by similar enzymes that are used for chain assembly, and the chemistry carried out during the synthesis has the same nature [4]. Fatty acid synthesis could be described as a special case of polyketide synthesis, where the resulting polyketide is always fully reduced to an aliphatic chain. In addition, fatty acid synthesis has been studied for longer, and it is therefore a good basis for study to understand polyketide synthesis.

1.3 The fatty acid synthesis

1.3.1 FAS biosynthetic pathway

Fatty acid biosynthetic pathway (Figure 1.2) consists of an iterative decarboxylative Claisen condensation between a starter unit and an extender unit to form a β-keto thioester. An acetate unit is firstly transferred from acetyl-CoA to the acyl carrier protein (ACP) by the malonyl-acetyl transferase (MAT), then to the active site of the condensing enzyme, the ketosynthase (KS). A malonyl extender unit is transferred from malonyl-CoA to the ACP by MAT, and then the condensation reaction occurring between both the starter and extender units produces an extended chain that is bound to the ACP. The chain is then reduced by a ketoreductase (KR), dehydrated by a dehydratase (DH), and reduced again by an enoyl reductase (ER). Finally, the nascent chain is transferred to the condensing enzyme and the cycle is repeated several times, until the chain is passed to a thioesterase (TE), which releases a free acid or acyl ester [5]. Usually, the chain length reaches a maximum

of 16-18 carbon atoms, yielding palmitoyl or stearoyl derivates [6]. Further extension cycles beyond carbon C18 can occur, but the ultimate length rarely exceeds 24 carbons [7].

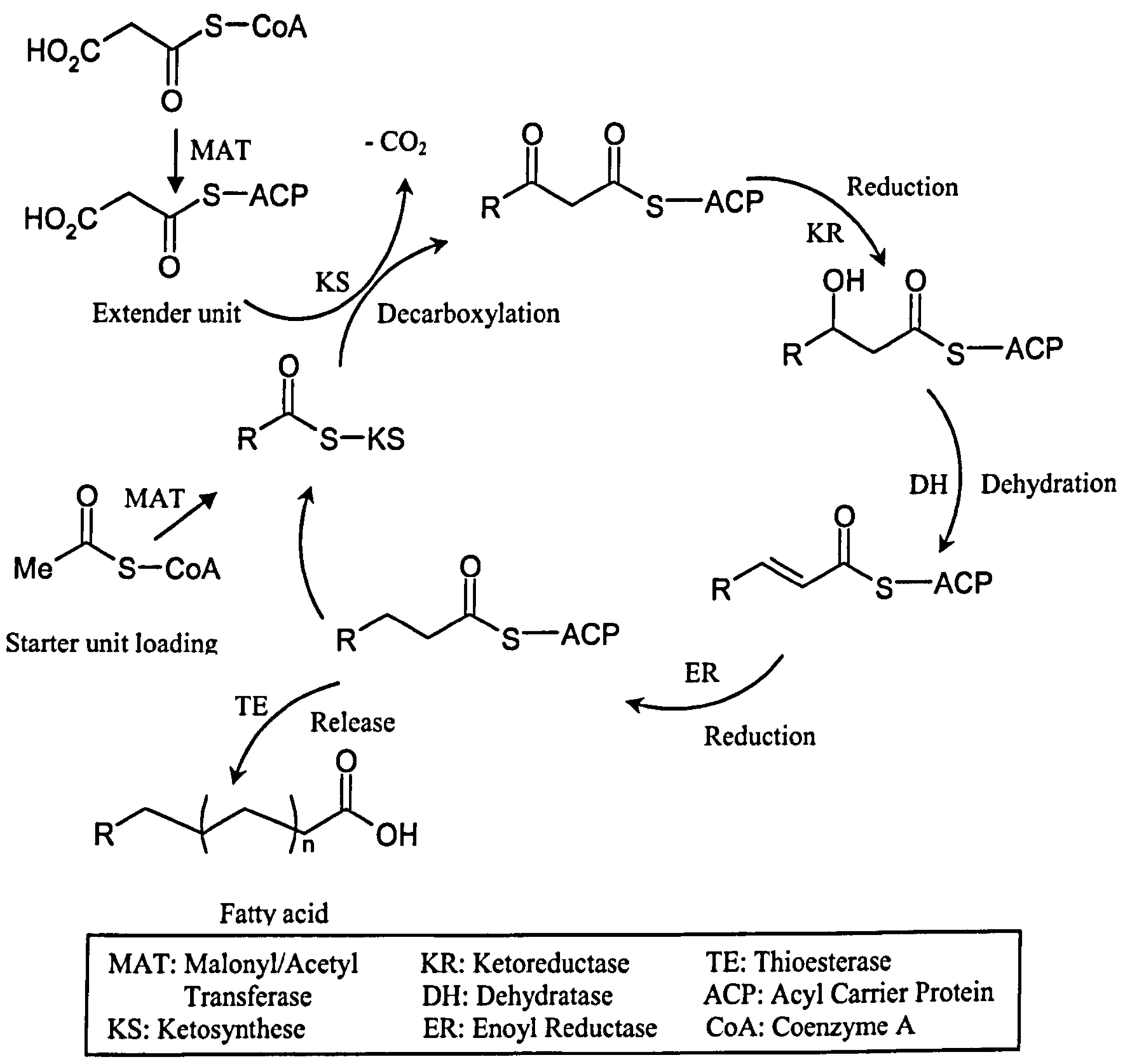


Figure 1.2 Fatty acid biosynthesis.

In this cycle, a starter acyl unit, usually acetyl, is condensed with a malonyl unit, which undergoes decarboxylation in order to provide the electrons for a new carbon-carbon double bond. These acyl participants are bound as thioesters to the enzymes involved in the chain extension. If the starter acyl is bound to a cysteine thiol of the ketosynthase, the condensing enzyme, the extender units are bound to a thiol that is not part of the primary sequence of the acyl carrier protein, but it is the terminus of the phosphopantetheine arm that is added to the ACP after post

translational modification. The organization of the different activities that characterize the fatty acid synthesis depends on the organism. Type II bacterial and plant fatty acid synthases are discrete proteins that can be isolated separately and reactions are carried out by these individual enzymes. On the other hand, type I animal synthases consist of different domains, each of them with a separate activity, linked covalently to form a large multifunctional enzyme [6].

1.3.2 Structure of type I mammalian FAS

Structurally, mammalian type I FAS consists of two identical polypeptide chains organised as dimers, forming two active sites at the homodimer interface [8, 9]. A model for the FAS organization was proposed, and it was based on a logical interpretation of the possible structure of a FAS monomer [10]. Observing that the bifunctional dibromopropane could cross-link the thiol of the N-terminal ketosynthase of one monomer with the C-terminal thiol of the ACP of another monomer, a model with two extended FAS monomers disposed in a head to tail antiparallel arrangement was proposed. The re-examination of the dibromopropane experiment [11] and functional complementation studies of baculovirus mutants [12] revealed new intra- and intermonomer interactions. Following these investigations, an alternative model for the FAS, in which the two monomers are intertwined and head to head contacts are formed (Figure 1.3), was proposed [13]. In particular, the two ACP domains can interact with the MAT or KS domains of either subunit. It was also proposed that the ACP interacts with MAT and KS of either subunit during substrate loading and the condensation reaction, while they interact only with one subunit during substrate processing and release [9]. It was also proved that a single wild-type subunit of the synthase, oligomerized with a mutant subunit where all the active sites have been destroyed by mutations, was still capable of synthesis, suggesting that, while the fatty acid synthase functions most efficiently when both the inter and intramolecular interactions are conserved, the synthesis can also occur when only one subunit is active [13]. Very recently, the X-ray structure of the mammalian FAS has been published at a resolution of 4.5 Å [14]. Whilst it was not possible to identify individual residues or trace the polypeptide backbone, but secondary structure elements were identifiable. The overall structure is very similar to the EM reconstruction [15], and appears to be

asymmetric in clear contrast with the antiparallel model (Figure 1.4). The central portion of the FAS structure is composed by the ketosynthase (KS) domain, which is dimeric, and by the dehydratase (DH) and enoyl-reductase (ER) domains. As well as the KS, the ER domain is dimeric, while the DH domains adopt a pseudo-dimeric fold within each monomer. The malonyl-acetyl transferase (MAT) and ketoreductase (KR) domains protrude from this central portion. Both MAT and KR domains assume a pseudo-symmetric arrangement and are monomeric. The location of the ACP and thioesterase domains was not clearly identified.

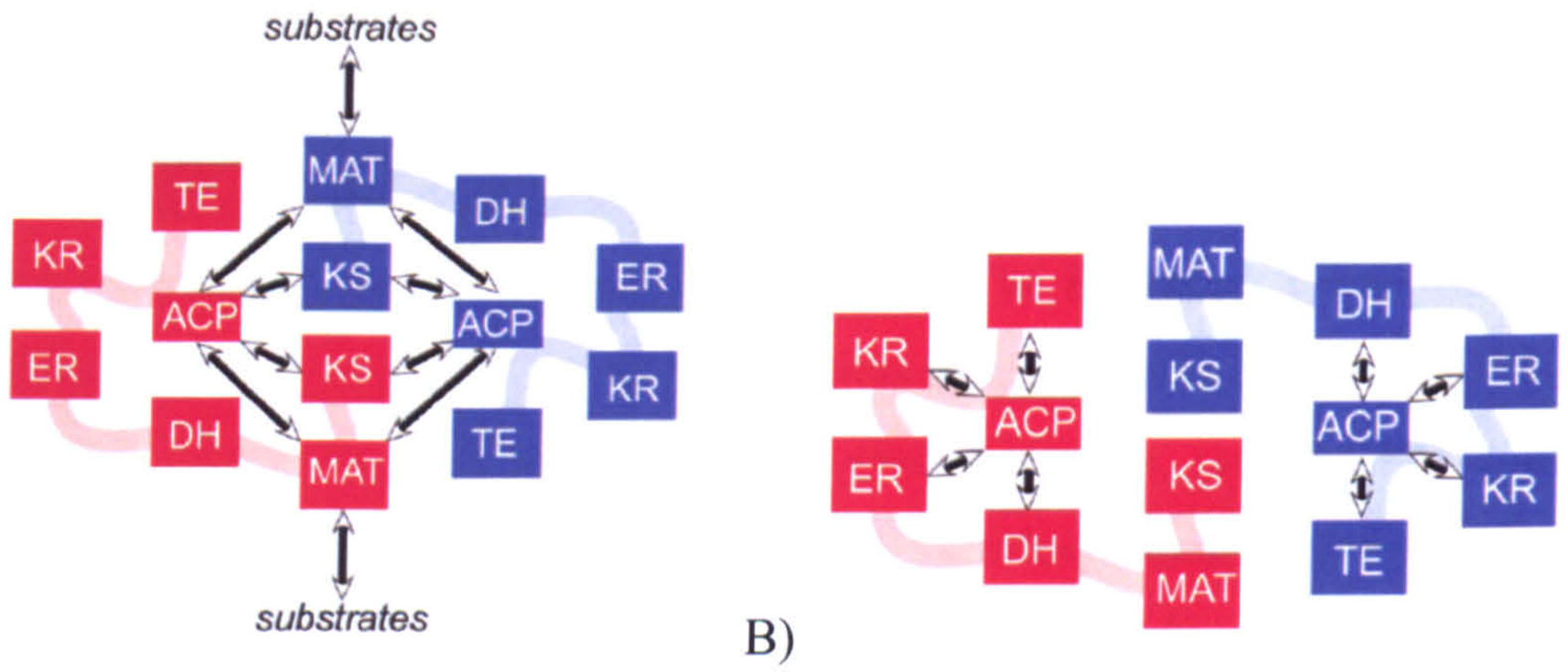


Figure 1.3 Model of the functional contacts between FAS domains. During substrate loading and condensation (A), ACP interacts with the ketosynthase (KS) and the malonyl/acetyl transferase of either subunit, while during substrate processing and release ACP interacts with the enzymes of the same subunit (Figure adapted from [9]).

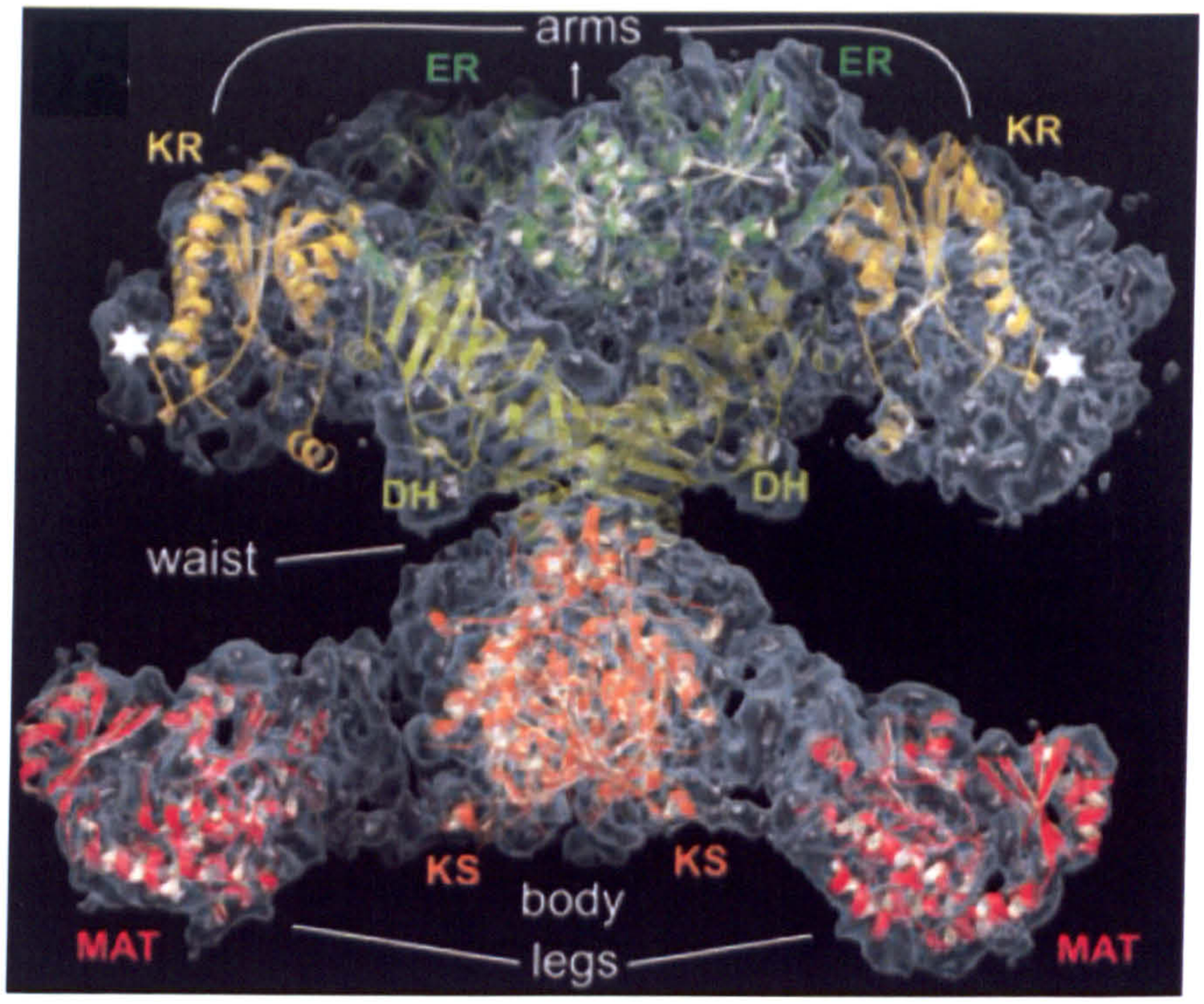


Figure 1.4 Organisation of the FAS active sites. The KS dimer forms the bottom part of the central portion, while MAT domains protrude from the bottom. DH and ER domains are located at the top of the central portion, and the KR domains extend from the top. White stars indicate the pseudosymmetry-related suggested attachment regions for ACP and TE (Figure adapted from [14]).

1.3.3 Fungal type I FAS

The architecture of vertebrates and fungal type I FAS are different. In fungi, the functional domains are organized in a 2.6-MDa barrel-shaped $\alpha_6\beta_6$ dodecameric complex [16]. Two polyfunctional proteins, α and β, assembly the fungal fatty acid synthase complex. In *Saccharomyces cerevisiae*, the 210 kDa α-chain contains a KS domain, a ketoacyl reductase and a phosphopantetheinyl transferase domain. In addition, it includes an N-terminal ACP, required for the transfer of intermediates between the different catalytic sites. The 230 kDa β-chain contains an acetyl transferase and a malonyl/palmitoyl transferase domain, a dehydratase and an enoyl reductase domain. The structure for the *S. cerevisiae* FAS was recently solved [17-19]. In addition, Ban *et al.* presented an atomic model of the *Thermomyces lanuginosus* type I FAS at 3.1 Å [20]. The organization of the *S. cerevisiae* FAS is illustrated in Figure 1.5. The barrel-shaped FAS complex is separated into two halves by a wheel-like structure at the equator formed by the α subunits. The dome

is formed by a trimer of β subunits which lie on each side of the α subunit wheel. The ACP domain is located inside the hollow particle and presumably moves within a limited space, which is called the reaction chamber [19].

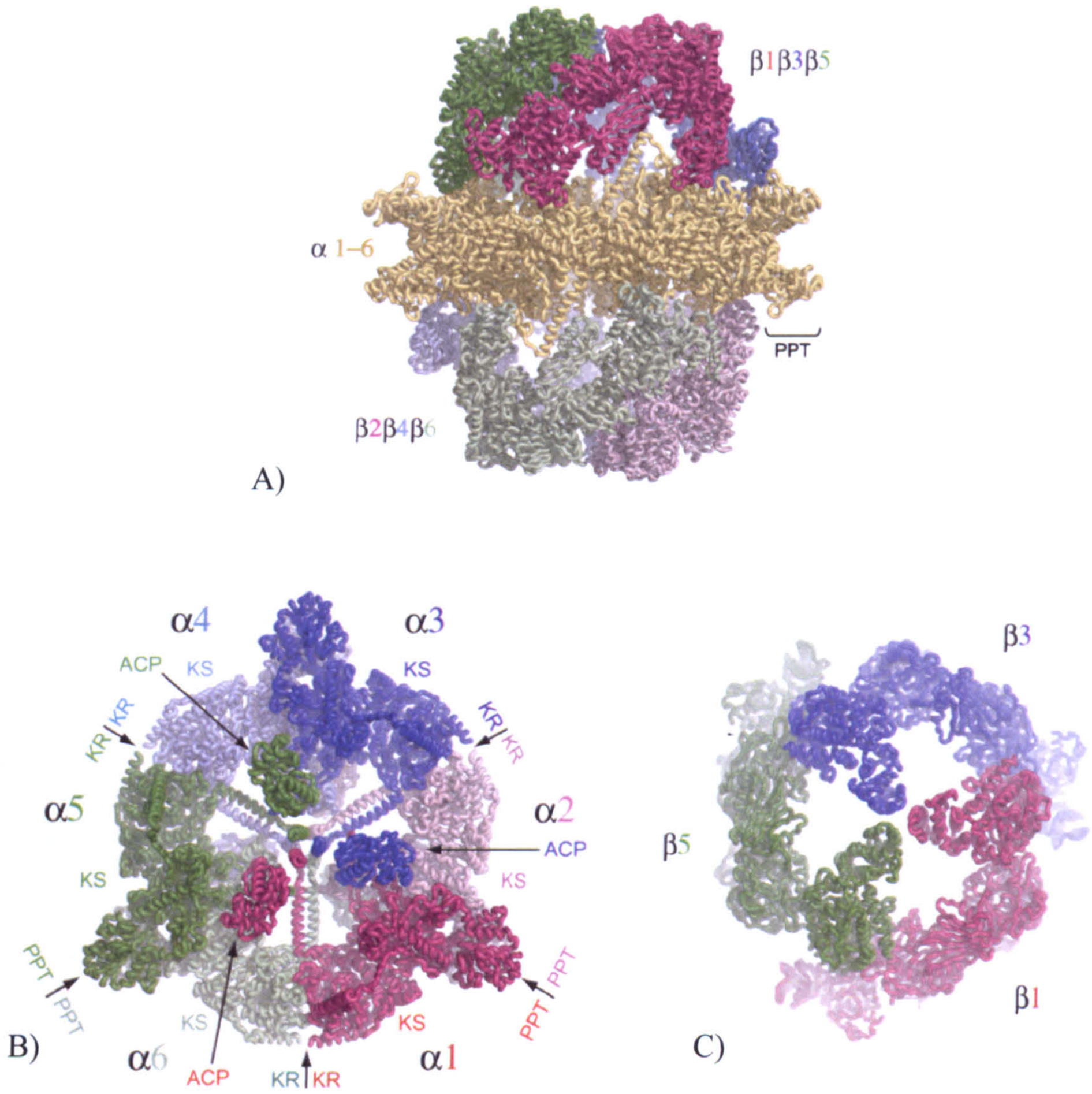


Figure 1.5 The *S. cerevisiae* type I FAS. A) The overall $\alpha_6\beta_6$ dodecameric complex. B) The top view of the α subunit. C) The β subunit trimer viewed from the top. Figure adapted from [19].

1.3.4 *E. coli* type II FAS

In the *E. coli* type II FAS, each reaction is carried out by a discrete enzyme, encoded by a unique open reading frame. Malonyl-CoA:ACP transacylase (FabD, MCAT) catalyses the malonate transfer from malonyl-CoA to the acyl carrier protein, a step that is essential for the beginning of the reaction but also for each

elongation cycle. The crystal structure of FabD [21, 22] revealed that this enzyme shares the same catalytic Ser-His dyad of serine hydrolases. The β-ketoacyl-ACP synthase III (FabH), the condensing enzyme, starts the reaction using CoA thioesters as primers, while the elongation condensing enzymes FabB and FabF utilize acyl-ACPs [23]. FabG, a member of the short chain reductase/dehydrogenase superfamily [24], is involved in the following NADPH dependent reduction of the β-ketoacyl-ACP. Analysis of the crystal structures of the plant *Brassica napas* FabG-NADPH complex and the apo protein from *E. coli* revealed that a conformational change occurs when the cofactor is bound, and ACP loop-2 region is associated to FabG recognition and binding [25]. FabA and FabZ, which vary only in their substrate specificities, catalyze the dehydration of the β-hydroxyacyl-ACP, producing the trans-2-enoyl-ACP [26]. Finally, the NADH-dependent reductase FabI reduces the enoyl-ACP [27, 28]. The following elongation steps are then catalyzed by the β-ketoacyl-ACP synthase I (FabB) and II (FabF), which are characterized by a His-His-Cys active site motif common to the thiolases.

1.4 Polyketide biosynthesis

Polyketide synthesis is very similar to fatty acid synthesis. It involves successive decarboxylative condensation cycles between simple acyl units to form a β-keto ester (Figure 1.6).

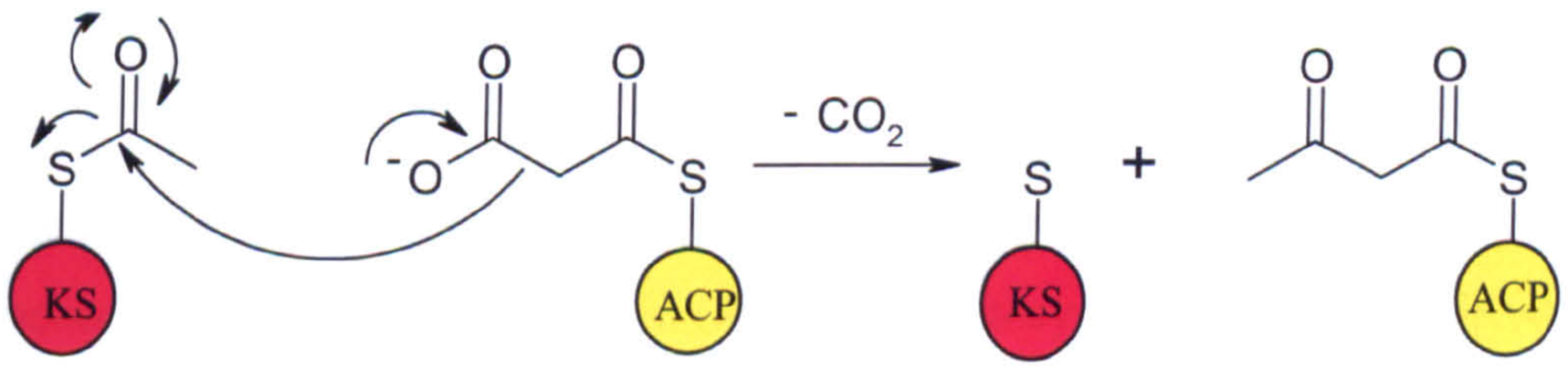


Figure 1.6 Claisen condensation between the starter and the extender units. The acyl starter unit is bound to the ketosynthase, while the malonyl extender unit to the ACP.

In contrast to fatty acid synthesis, polyketide synthesis involves variable cycles, with more complex modifications, and it needs to be more programmed than the fatty acid synthesis. The choice of the building units, their order and stereochemistry, and the order of further modifications are predetermined. Moreover, the polyketide chains are in many cases cyclised to produce macrocyclic rings, and the presence of multiple chiral centres and other modifications, for example the addition of sugars, methyl and hydroxyl groups, increase their structural complexity. In comparison to the fatty acid synthesis, a wider range of acyl starter units and modifying enzymes, including cyclase, oxygenases, NADPH and FADH- dependent oxidoreductases, acyl transferases, glycosyltransferases and *N*-methyl transferases, are present [29] (Figure 1.7).

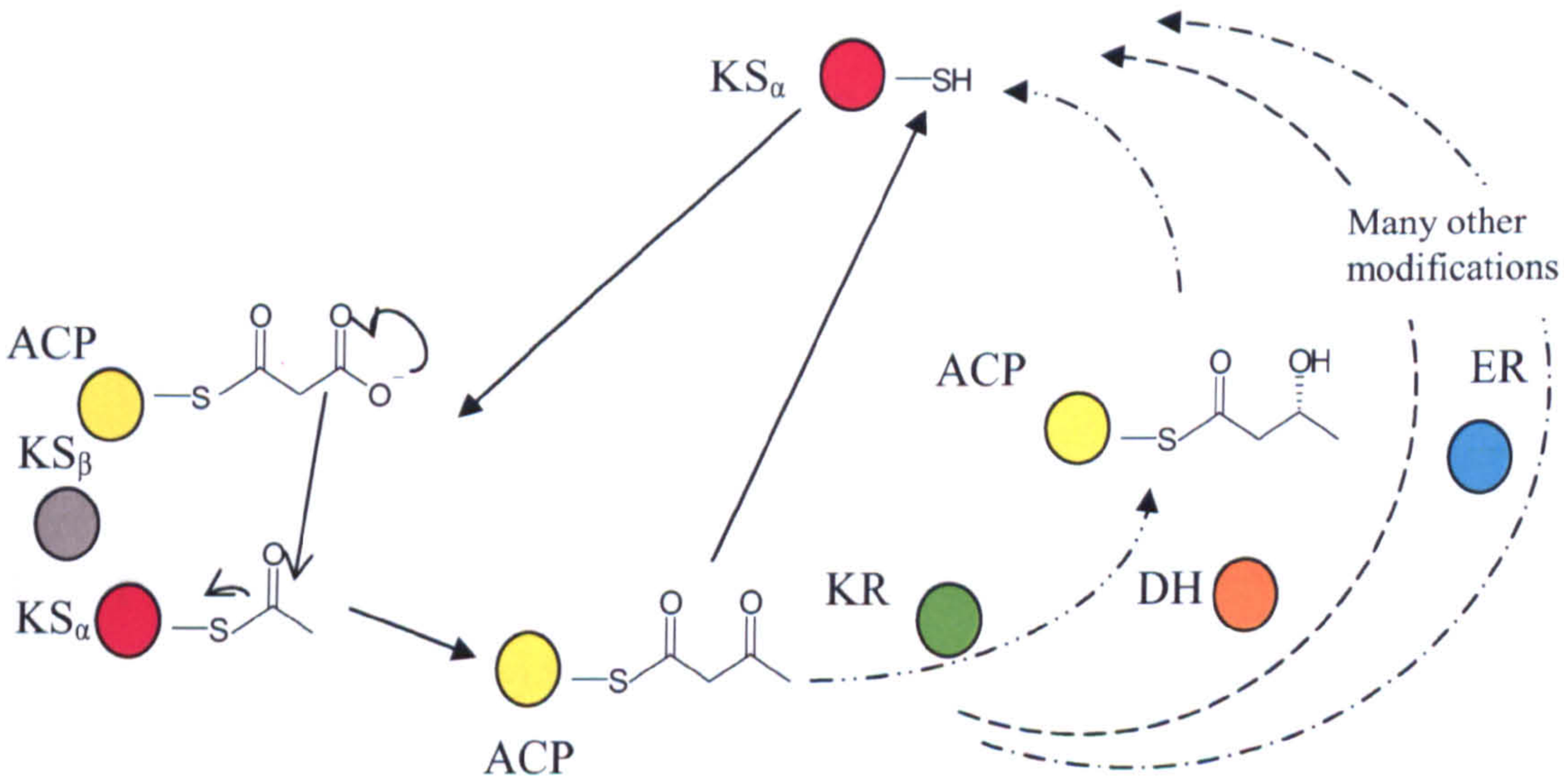


Figure 1.7 Polyketide synthesis. Polyketide synthesis involves a minimal system, composed of the acyl carrier protein (ACP), the ketosynthase KS_α and KS_β (called also chain length factor (CLF)). This complex catalyses the condensation reaction, while other enzymes, such as the ketoreductase (KR), the dehydratase (DH) or the enoyl reductase (ER) can modify the growing chain before the beginning of a new extension cycle.

1.4.1 Polyketide synthases

Each polyketide is synthesized by its specific polyketide synthase (PKS). Three types of polyketide synthases can be identified on the basis of their structure and mechanism. Type I PKSs, such as the erythromycin synthase, are giant multienzymatic proteins carrying all the active sites required for the biosynthesis of reduced polyketides, and they can be modular or iterative type I PKS. Type II PKSs, for example the one responsible for the production of the actinorhodin from *Streptomyces coelicolor*, are made up from single mono-functional enzymes that catalyse the formation of aromatic compounds, but not extensively reduced. These PKSs are analogous to the bacterial fatty acid synthases and are involved in the biosynthesis of bacterial aromatic natural products such as actinorhodin, tetracenomycin and doxorubicin. Finally, bacterial or plant type III PKSs, such as the stilbene or chalcone synthases, are usually single peptides that behave like condensation enzymes, without using ACP domains and acting directly on the coenzyme A thioesters of simple carboxylic acids [30]. While type I and II PKSs are structurally different but share a high level of sequence similarity, type III PKSs act as condensing enzyme and show little sequence similarity to the ketosynthases of either type I or II PKSs [31-33].

1.4.2 Type I PKSs

Type I PKSs are multimodular domains similar to the type I fatty acid synthase. This group can be subdivided in iterative type I PKSs, such as the lovastatin nonaketide synthase (LNKS) [34], and modular PKSs, for example the 6-deoxyerythronolide B synthase (DEBS).

1.4.2.1 Iterative type I PKSs

Iterative Type I synthases are commonly found in fungi, and consist of a single module that is used iteratively, so that at least some of the constituent enzymes act in every cycle. Examples are the synthases involved in the biosynthesis of the 6-methylsalicylic acid (6-MSA) and lovastatin.

1.4.2.1.1 6-Methylsalicylic acid synthase

6-MSA is assembled from one molecule of acetate and 3 of malonate (Figure 1.8). The PKS responsible for its synthesis consist of KS, MAT, DH, KR and ACP (in this order). These domains act repeatedly catalysing three rounds of chain extension, with different levels of reduction at each stage, using a mechanism that is not known. For example, while the first condensation is followed by reaction with a second malonate, the second condensation is followed by reduction and dehydratation of the new keto group. After the third cycle, the polyketide chain is cyclised, dehydrated and enolised to yield the aromatic 6-MSA [35] A thioesterase is not included in this system, probably because the release of 6-MSA does not occur by hydrolysis.

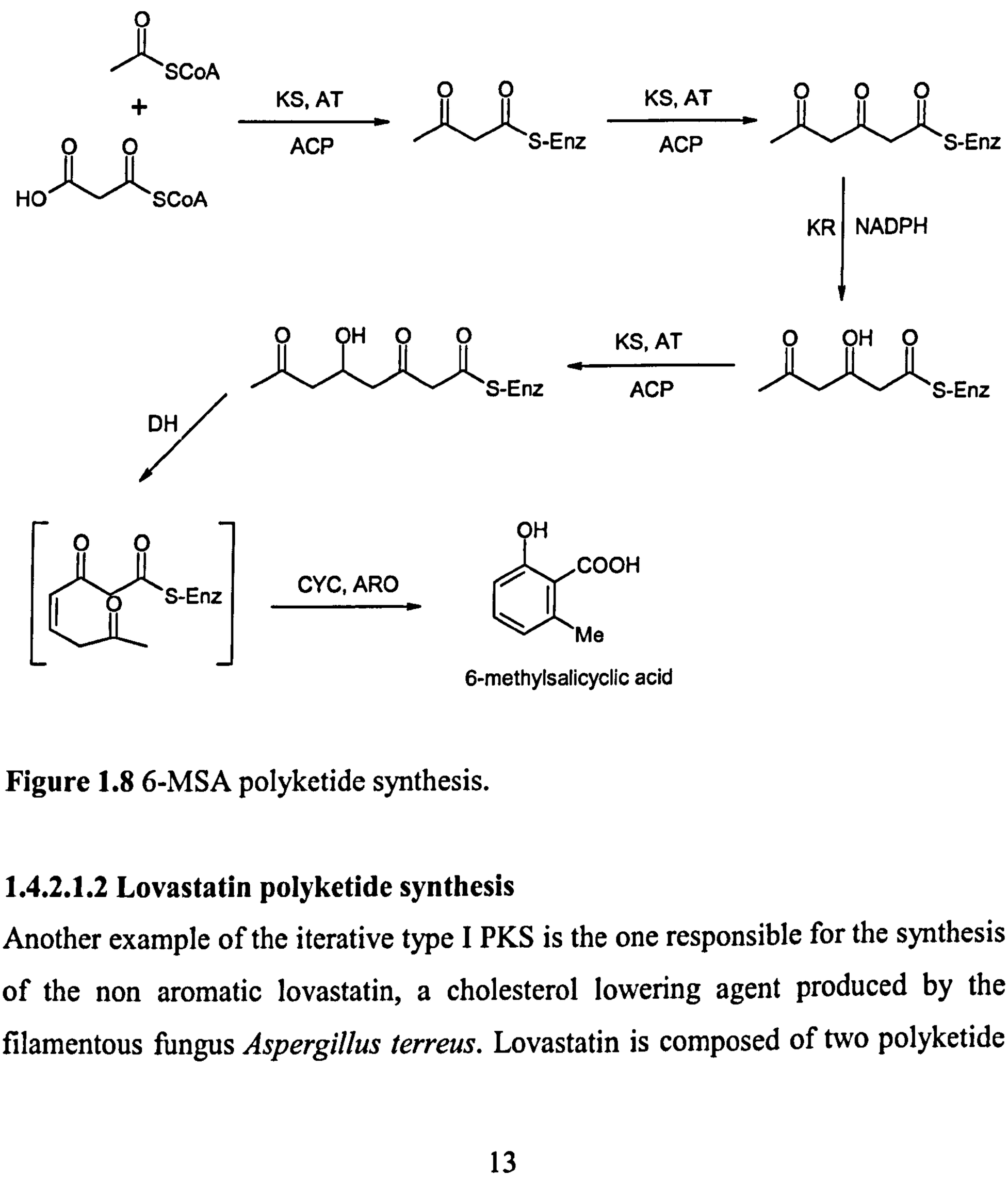


Figure 1.8 6-MSA polyketide synthesis.

1.4.2.1.2 Lovastatin polyketide synthesis

Another example of the iterative type I PKS is the one responsible for the synthesis of the non aromatic lovastatin, a cholesterol lowering agent produced by the filamentous fungus *Aspergillus terreus*. Lovastatin is composed of two polyketide

chains joined via an ester linkage [36]: one chain is a nonaketide, synthesized by the lovastatin nonaketide synthase (LNKS), the other a diketide produced by the lovastatin diketide synthase (LDKS). Both PKSs generate their product using an acetate starter unit and malonate extender units, adding also a methyl group from S-adenosyl methionine (Figure 1.9). The LNKS is unimodular and contains KS, AT, DH, ER, KR and ACP domains in the same order as in the animal FAS, with the additional presence of a methyl transferase (MeT) and the substitution of the TE domain with a putative peptide synthetase elongation domain (PSED). It is believed that LNKS works in cooperation with the discrete type II protein LovC (a putative ER), to synthesize dihydromonacolin L [37]. Dihydromonacolin L is then tailored to monacolin J. LDKS, which produces (2R)-2-methylbutyrate, is also unimodular with the same organization of LNKS, but without the PSED domain and with an active ER. The synthesis of lovastatin is completed by the attachment of the (2R)-2-methylbutyrate to monacolin J by a specific transesterase enzyme.

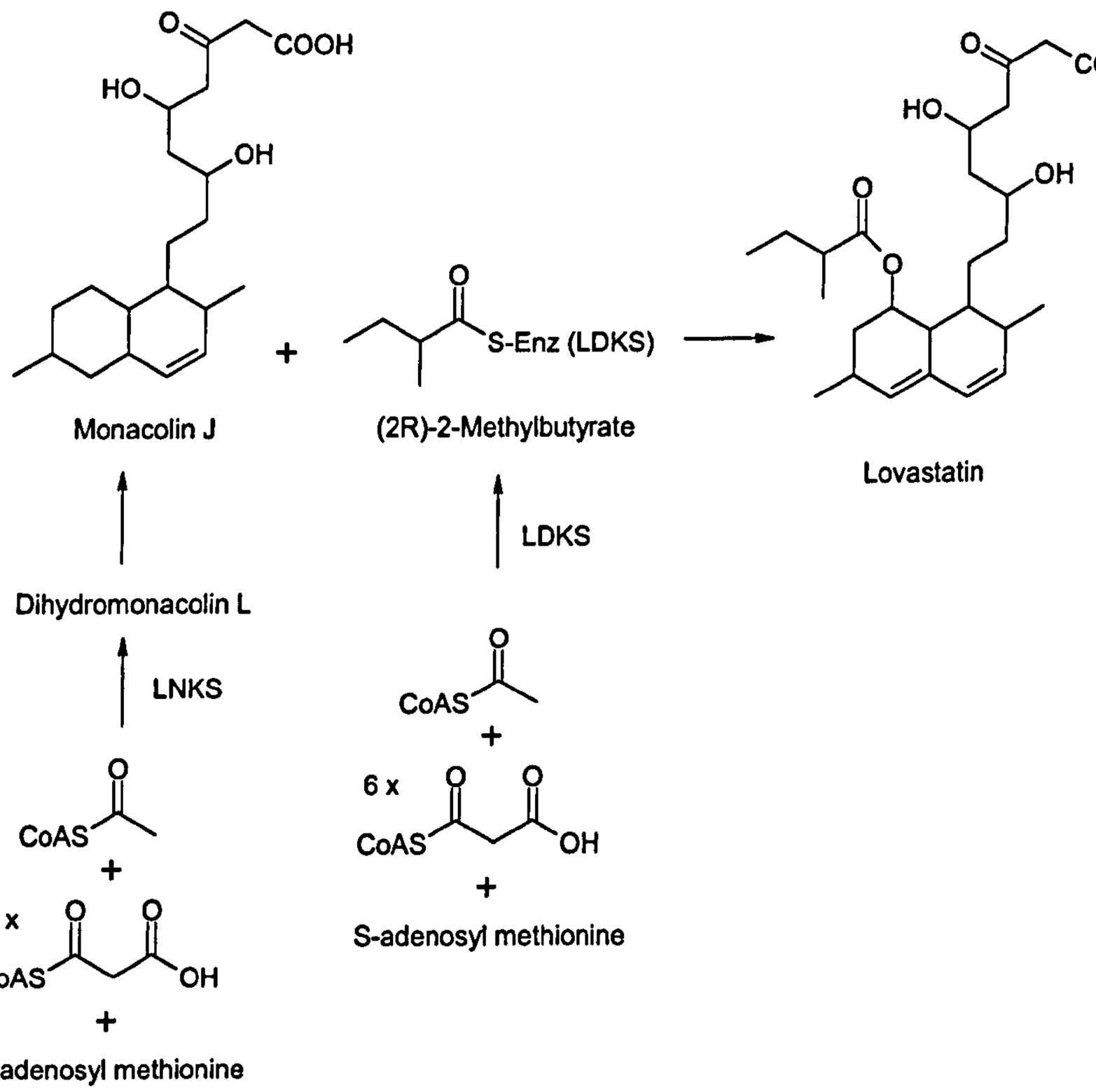


Figure 1.9 Lovastatin polyketide synthesis. The nonaketide and diketide portions are generated by two independent polyketide synthases.

1.4.2.1.3 SQTKS

Polyketides produced by fungi show a wide diversity of both structural motif and activity. In some fungal PKSs, such as *Aspergillus oryzae*, a methyltransferase (MeT) domain was also identified. Squalestatin S1 (also known as zaragozic acid A) is a powerful inhibitor of squalene Synthase, identified as a metabolite of *Phoma sp.* and other filamentous fungi [38]. Squalestatin can be used in the treatment of diseases associated with a high level of serum cholesterol concentrations, and it is composed of 2 polyketide chains (Figure 1.10). The main chain is a hexaketide (chain A), composed of a benzoate derived starter unit and five acetate-derivate extender unit, while the side chain (chain B) is a tetraketide, with acetate as starter unit. Four other carbons are derived from succinate, forming the acid core.

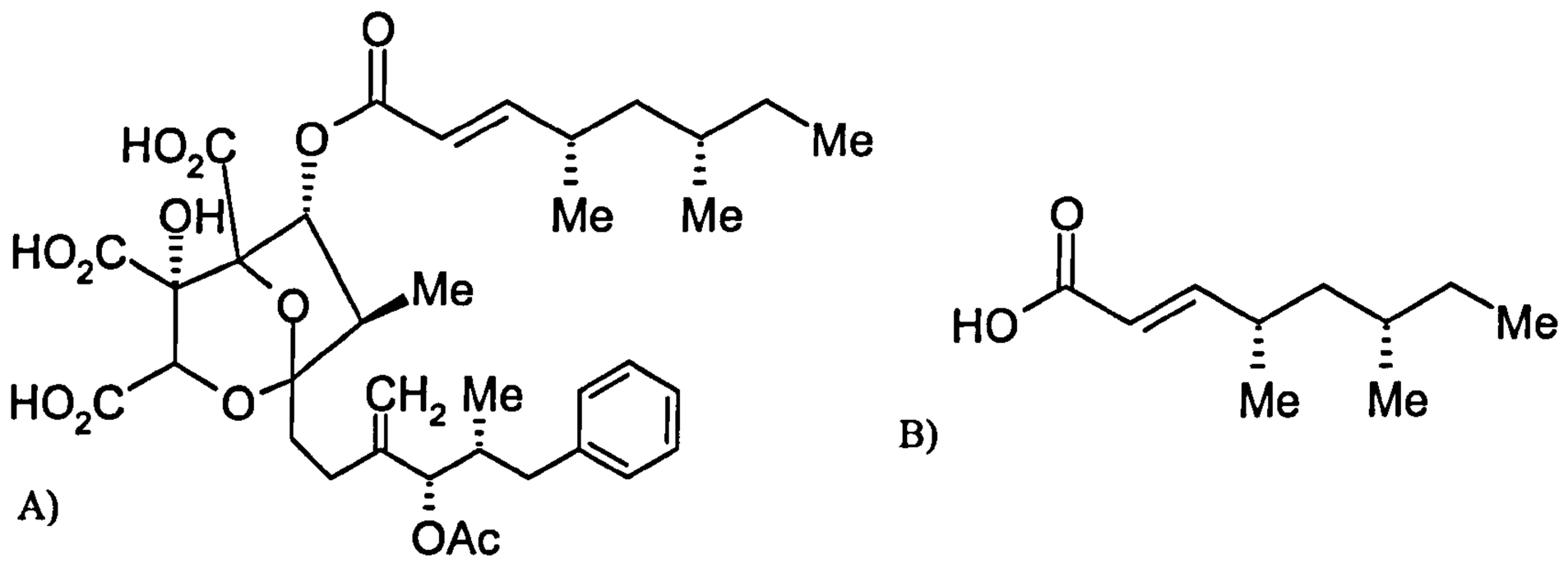


Figure 1.10 Squalestatin S1 (A) and its chain B (B).

After several experiments, including the generation of cDNA and PCRs, a 8142 bp PKS (Phoma PKS1) contained KS, AT, DH, C-MeT, ER, KR and ACP domain encoding regions, was inserted into the expression vector pTAex3 forming a 15747 bp expression construct [39]. This vector utilizes an *amy*B promoter, which is induced by starch or maltose and is repressed by glucose. The translated gene encodes a polypeptide of 2714 amino acids. After expression in *Aspergillus oryzae*, organic extraction and RP-HPLC analysis, a new compound was found to be expressed. The characterization revealed that it was the B chain of squalestatin, and the protein encoded by Phoma PKS1 was named SQTKS (Squalestatin Tetraketide Synthase). It seems that all the PKS1 domains are active. The biosynthesis must

involve 3 rounds of chain extension, with the MeT occurring after the first and second round, and KR and DH active in all rounds. ER and C-Met are not active in the final round. SQTKS so can be a great starting point for studying fungal type I PKSs. It is important to point out that the auxiliary domains are involved selectively in specific cycles during the polyketide production.

1.4.2.2 Modular type I PKSs

Modular PKS, for example the systems responsible for the synthesis of macrolides like erythromycin, rapamycin and rifamycin, consists of multiple active domains organised into modules. Each module is responsible for the decarboxylative condensation reaction, followed by a programmed reductive cycle. These modules are active only once during the production of the polyketide, and usually in a sequential way. In addition, a loading module carries the starter unit at the front of the first module and a thioesterase unloads the polyketide at the end of the last module.

1.4.2.2.1 The erythromycin polyketide synthase

One example of modular type I PKS is given by the synthase responsible for the synthesis of the polyketide portion of the antibiotic erythromycin, the 6-deoxyerythronolide B (6-dEB). 6-dEB is produced by sequential condensation of propionyl-CoA with 6 units of methylmalonyl-CoA as enzyme bound intermediates, followed by lactonisation to 6-dEB [40-42]. Three enormous open reading frames, responsible for the production of 6-dEB: eryAI, eryAII and eryAIII, encode for three huge multienzyme polypeptides, 6-deoxyerythronolide B synthase (DEBS) 1, 2, and 3 (Figure 1.11). Linker regions between each domain are vital to maintain the correct topology. Each DEBS protein contains two functional modules, and each module is composed of the three essential catalytic domains, KS, AT and ACP, plus a variable set of other domains [43-45]. The three essential domains cooperate to catalyse the carbon-carbon bound by Claisen condensation, while the variable set of domains located between the AT and ACP (shown as a loop above the line of essentials domains in Figure 1.11) carry out the reductive modifications. No reductive domains are found in module 3, while the others incorporate a KR domain. Module 4 is characterized by a complete set of reductive

activities (DH, ER and KR). Structural studies of the DEBS proteins, including crosslinking experiments, have inspired a head to head homodimeric double helical structure [46] (Figure 1.12).

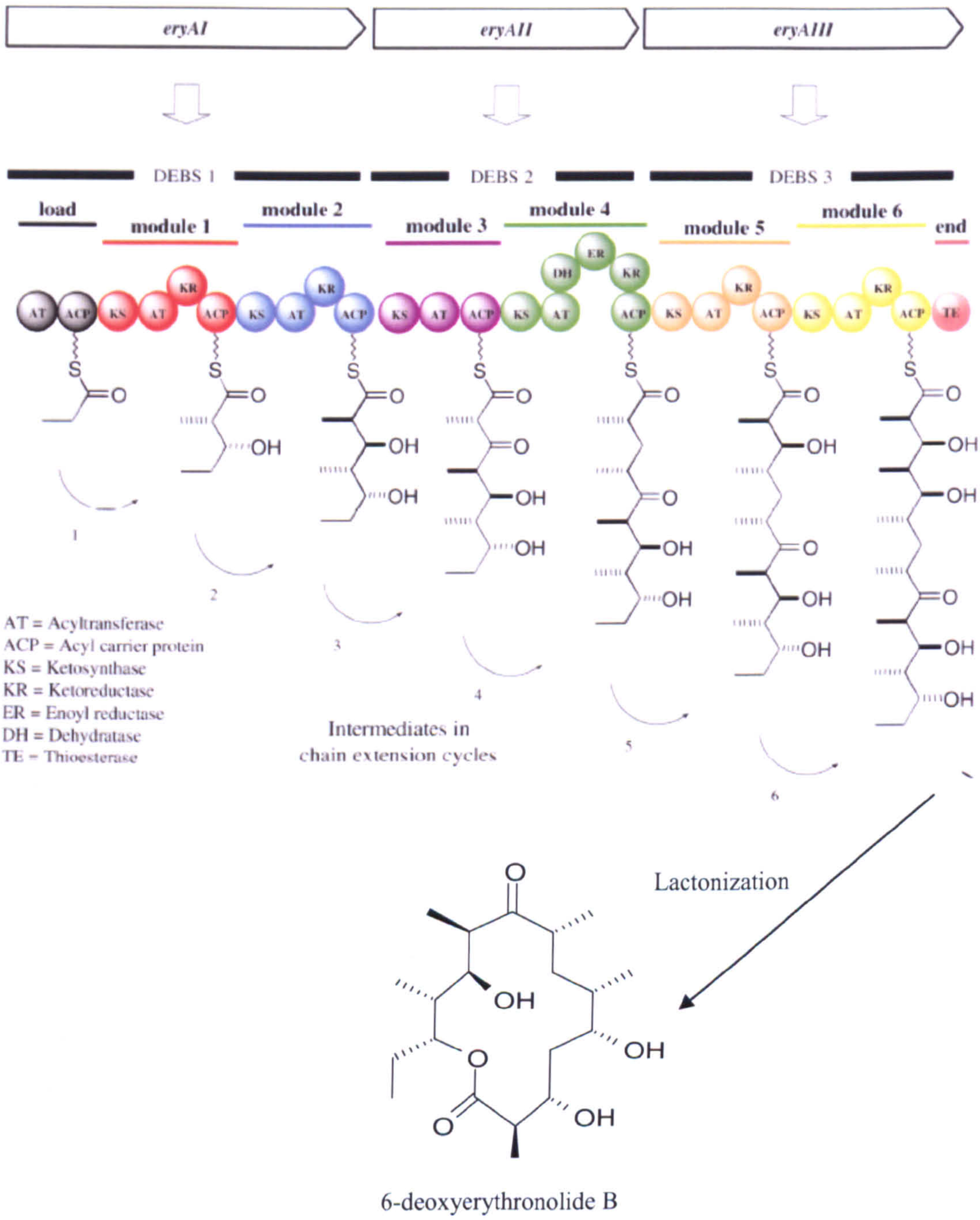


Figure 1.11 Domain organization of the erythromycin polyketide synthase. Each module incorporates the essential KS, AT and ACP domains, while all except one include optional reductive activities (Figure adapted from [3]).

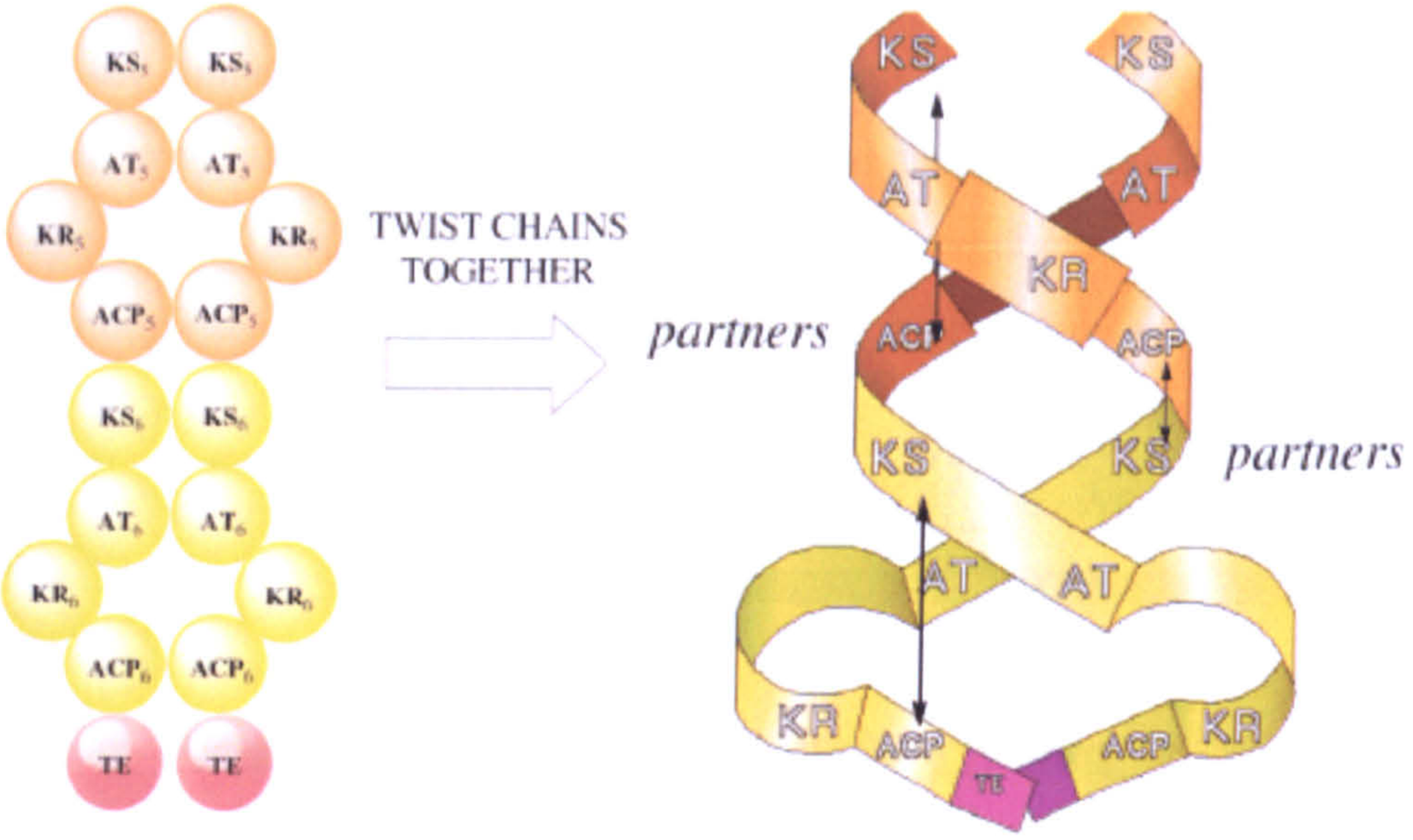


Figure 1.12 Double helical "Cambridge model" for the erythromycin synthase. The ACP can cooperate with KS and AT in the other module (Figure adapted from [3]).

1.4.3 Type II PKSs

Type II PKSs are discrete enzymes that can be used several times during the production of a polyketide [5, 47, 48]. The products of these synthases are generally phenolic aromatic compounds. Polyketides are produced by the condensation of building blocks, and subsequent cyclization and aromatization of these chains occur.

1.4.3.1 Type II PKS gene cluster organization

The most studied type II PKS is the synthase involved in the production of actinorhodin by *Streptomyces coelicolor*, a gram positive bacterium that produces the polyketide antibiotics actinorhodin and methylenomycin, the immunosuppressive undecylprodigiosin, the non-ribosomal peptide antibiotic CDA and also the usual range of bacterial fatty acids. The gene cluster responsible for the actinorhodin biosynthesis was identified in 1992 by Fernandez-Moreno [49]. The identification of other type II gene clusters suggests a similar gene organization and consequently a similar evolutionary pathway (Figure 1.13). By using gene deletion experiments [50-53], it was discovered that a minimal set of three enzymes, encoded by the actI gene, is required *in vivo* for the catalysis: the acyl carrier protein, the ketosynthase α, the condensing enzyme, and the ketosynthase β, also

known as chain length factor (CLF) or chain initiation factor (CIF) [54]. For example, the unnatural polyketides SEK 4 and SEK 4b, produced by the actinorhodin minimal PKS, are obtained after 8 condensation reactions followed by cyclization [53]. Then other enzymes, such as ketoreductases, aromatases and cyclases, modify the nascent polyketide to produce the final product.

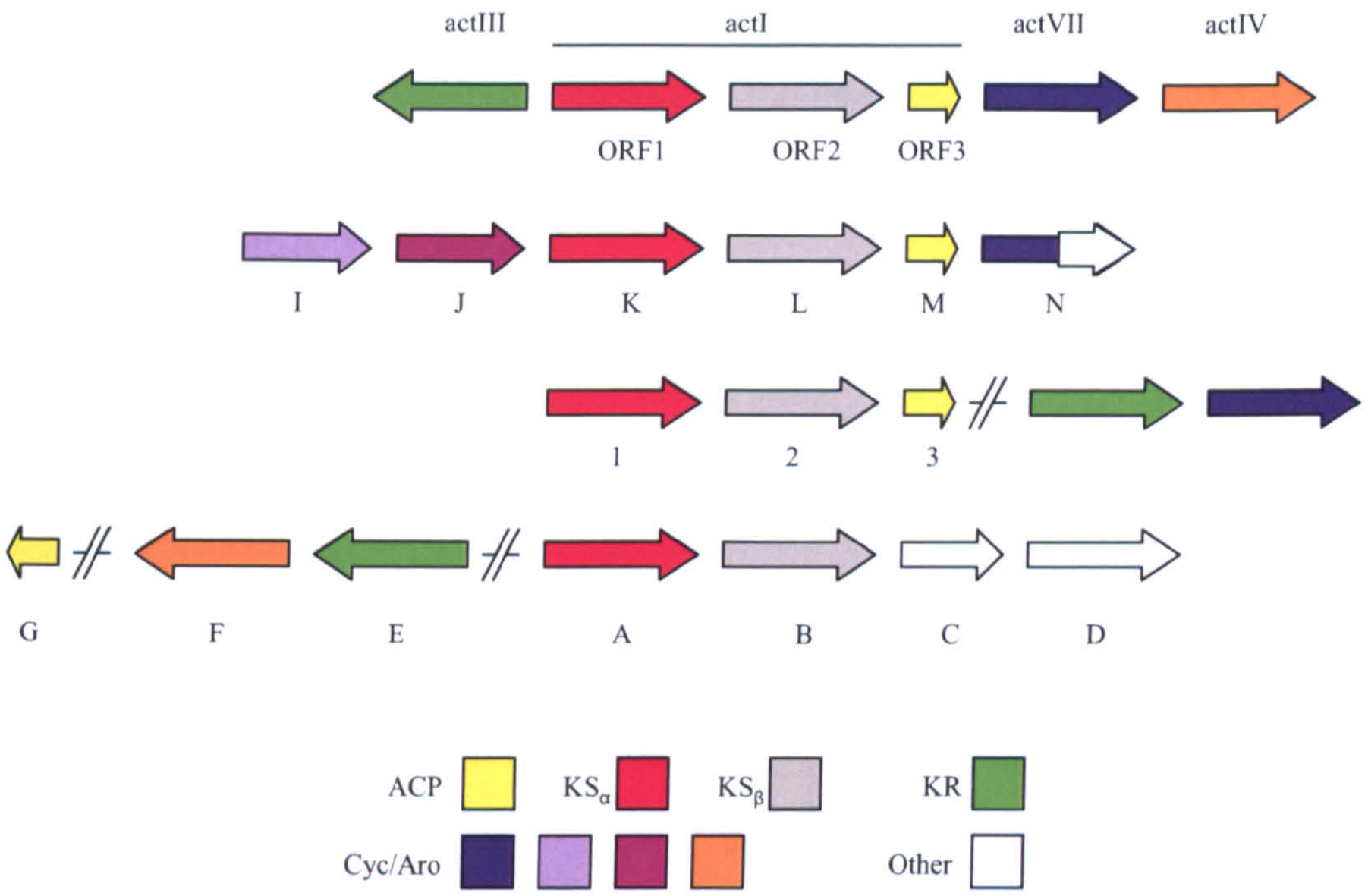


Figure 1.13 Organisation of gene cluster for type II PKSs. From the top: actinorhodin (act), tetracenomycin (tcm), oxytetracycline (otc) and daunorubicin (dps/dau). Figure adapted from [3].

1.4.3.2 Acyl Carrier Protein

The biosynthesis of polyketide secondary metabolites takes place on large multienzymatic complexes starting from small building blocks [5]. In both polyketide [4, 55, 56] and fatty acid [8, 57, 58] synthesis, and in the biosynthesis of non-ribosomal peptides [59], specialized acyl carrier proteins (ACPs) hold the assembly intermediates during biosynthesis. So, the acyl carrier proteins are very important during type I and II fatty acid and polyketide biosynthesis. Their phosphopantetheine acts as a long flexible arm that carries and delivers the nascent chain to the right destination. ACP is expressed as the inactive apo-protein, that is then converted to the holo form by a post- translational modification [60]. The 4'

phosphopantetheinyl (P-pant) moiety is transferred from the coenzyme A (CoA) to a conserved serine on the ACP. The side chain of this serine acts as a nucleophilic group attacking the pyrophosphate linkage of CoA. This modification is guided by the acyl carrier protein synthase (AcpS) in a magnesium dependent reaction [61] (Figure 1.14). The phosphopantetheine arm fulfils the two essential demands in fatty acid and polyketide biosynthesis: the intermediates remain covalently attached to the protein and the length and flexibility of the phosphopantetheine arm, approximately 20 Å, allows the transport of the intermediate between the different active sites of the set of enzymes involved in the catalysis.

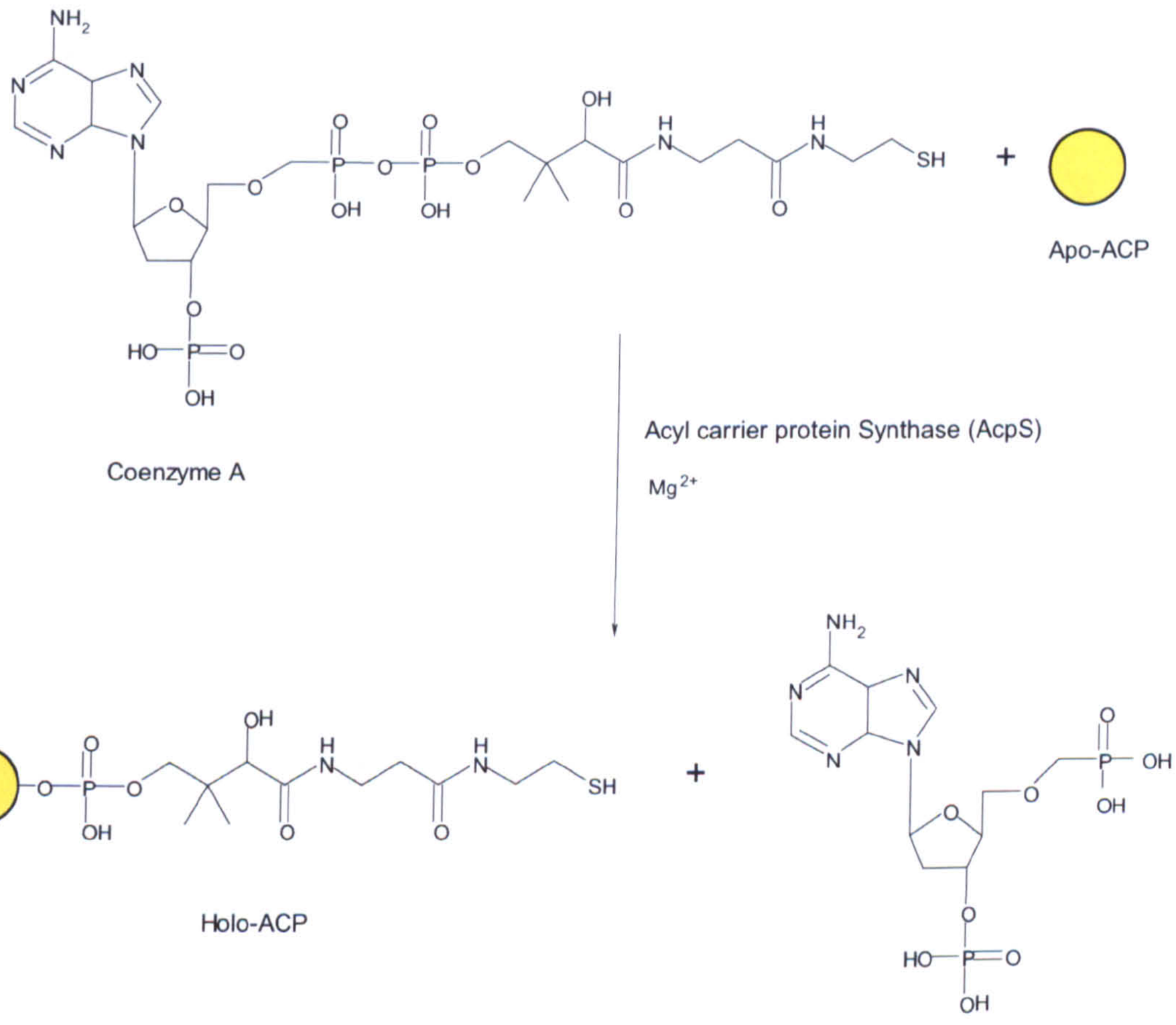
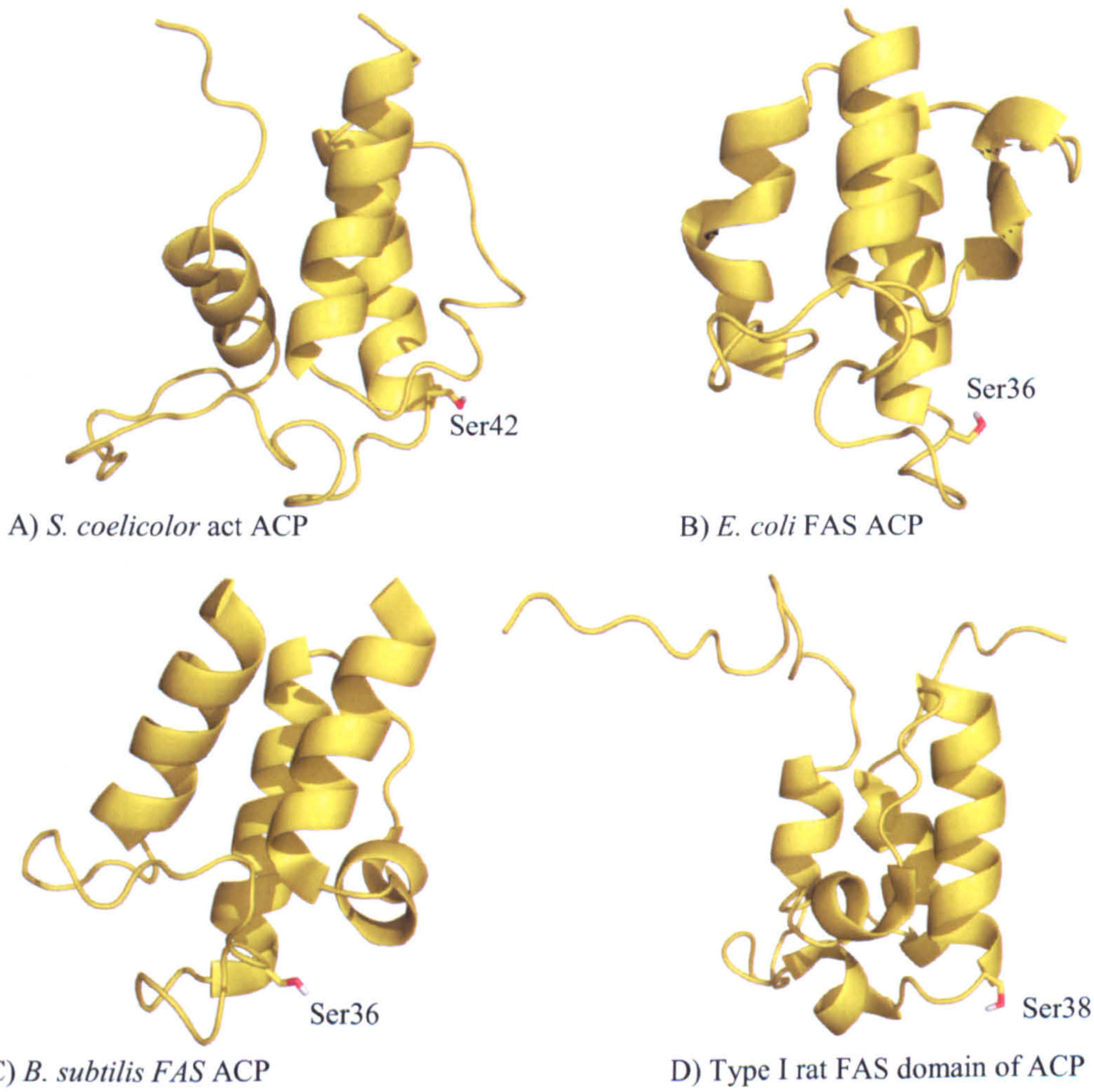


Figure 1.14 Phosphopantetheinylation of apo-ACP. The post-translational reaction is catalyzed by the Acyl carrier protein Synthase (AcpS) in the presence of CoA and magnesium.

ACPs share a high level of structural and amino acidic similarity. They can be described as a characteristic bundle of small negatively charged α-helices. Different ACPs, such as the *E. coli* [62] and *B. subtilis* [63] fatty acid ACPs, the actinorhodin polyketide ACP [64], the rat [65] and human [66] fatty acid ACP domain, and the oxytetracycline [67] and frenolicin [68] polyketide ACPs, have been solved by NMR and X-ray crystallography (Figure 1.15). Actinorhodin ACP from *Streptomyces coelicolor* was the first polyketide synthase ACP to be solved by 1D and 2D NMR. It is a small protein composed of 86 amino acids, and it consists of three major helices, a shorter helix, and a large loop region that separates helices 1 and 2. The site of the phosphopantetheinylation, Ser42, is located at the base of the negatively charged helix α2 [64, 69].



A) *S. coelicolor* act ACP

B) *E. coli* FAS ACP

C) *B. subtilis FAS* ACP

D) Type I rat FAS domain of ACP

Figure 1.15 Structures of different apo-ACPs. Cartoon representation of: A) actinorhodin polyketide ACP (PDB code: 2af8); B) *E. coli* fatty acid ACP (1acp) C) *B. subtilis* fatty acid ACP (1hy8) D) Rat fatty acid domain of ACP (2png). The active serine is shown as sticks.

It was also proved that the C17S actinorhodin ACP mutant, expressed to prevent the formation of a disulphide bond between the phosphopantetheine thiol and Cys17, can undergo self malonylation *in vitro* in the absence of a malonyl-CoA:ACP transferase [70]. This self-malonylation activity of act ACP is in contrast with what reported by Khosla and Hutchison. MCAT recruited from the fatty acid synthesis seems to be essential for the biosynthesis of polyketides in the actinorhodin and tetracenomycin minimal polyketide synthases [47, 71]. However, more recent work carried out in the Department of Chemistry, Bristol, has shown that act ACP can self acylate with different unnatural starter units derived from β keto N-acetylcysteamine (NAC) thiolesters, that are not substrates for the MCAT, giving further evidence of the self-acylation activity of actinorhodin ACP [72].

1.4.3.3 Acyl carrier protein Synthase

The important phosphopantetheine arm of ACP is transferred from the coenzyme A (CoA) to a conserved serine residue of the ACP in a post-translational modification guided by the acyl carrier protein synthase (AcpS) [61, 73]. AcpS belongs to the phosphopantetheinyl transferase superfamily, which can be classified into three groups, according to their sequence homologies and substrate spectrum. Group I contains bacterial proteins of about 120 residues, that transfer the phosphopantetheine arm onto fatty acids ACPs, and have been shown to accept type II PKS ACPs [74, 75]. The enzymes responsible for the modification of the peptidyl carrier protein subunits of non-ribosomal peptide synthases, of which Sfp from *Bacillus subtilis* is an example [76], are members of group II, that are bigger proteins, usually composed of 240 amino acids. Sfp exhibits a broader substrate specificity than other type I AcpS, and it is able to modify fatty acid and polyketide ACPs [61, 77]. The crystal structure reveals a two-domain enzyme, where each domain has a fold which is similar to the one of type I AcpS subunit [78] (Figure 1.16). A mechanism was proposed whereby a water molecule involved in Mg^{2+} complexation will be replaced by the hydroxyl group of the serine residue of the substrate PCP. This hydroxyl group can be deprotonated and act as nucleophile for the β-phosphate of CoA. In order to prevent self-hydrolization, the deprotonating base must be displaced by conformational change or must be provided by the PCP substrate [78].

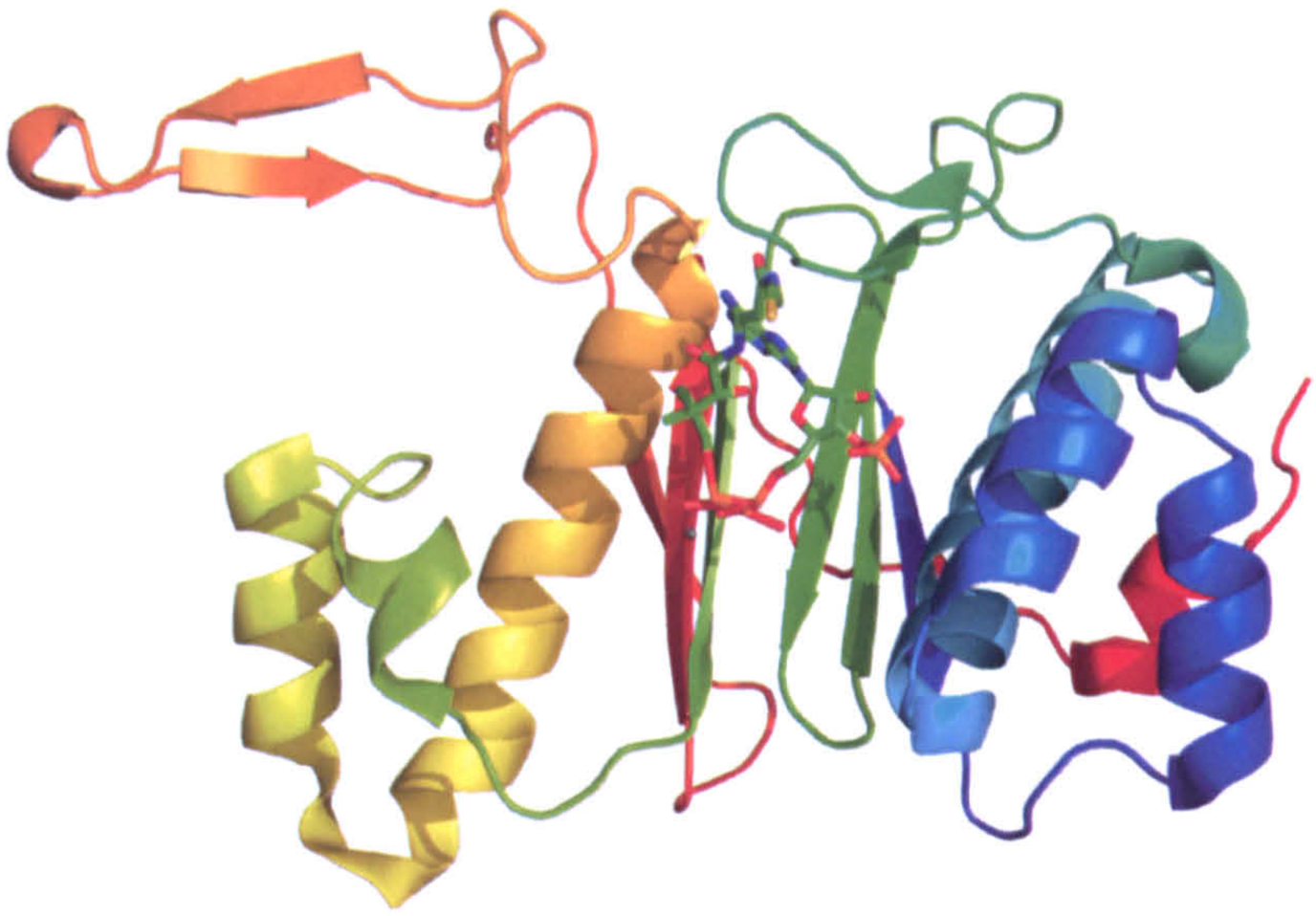


Figure 1.16 Surfactin syntethase-activating enzyme Sfp. Cartoon representation of the Sfp in complex with CoA, coloured in green (PDB code: 1qr0).

Group III includes proteins that are part of the heteromeric type I yeast and fungal FASs [79]. They seem to transfer the phosphopantetheine chain to their acceptor ACPs before the assembly of the protein complex they belong to.

Typically, organisms use one phosphopantetheinyl transferase for each individual phosphopantetheine-dependent pathway. For example, *Escherichia coli* has three phosphopantetheinyl transferases: the AcpS involved in the fatty acid synthesis, EntD that catalyses the synthesis of the siderophore enterobactin, and the product of the gene *yhhU*. Unlike Sfp, these AcpSs normally recognize their cognate ACPs [80].

The *E. coli* AcpS gene forms an operon with the upstream gene *pdxJ*, which function is required for the biosynthesis of vitamin B_6 [81, 82]. The AcpS gene was originally identified as *dpj* (downstream of *pdxJ*). The gene, essential for the growth of *E. coli*, encodes for a very low abundant small protein of about 14 kDa, and it was also shown that it can exhibit a discrete substrate specificity and can utilize a variety of ACPs [60, 75, 83-85].

Recently, the crystal structure of the acyl carrier protein synthase from *Bacillus subtilis* involved in fatty acid synthesis has been determined with and without the coenzyme CoA, and in complex with the *B. subtilis* type II holo-ACP [86] (Figure 1.17).

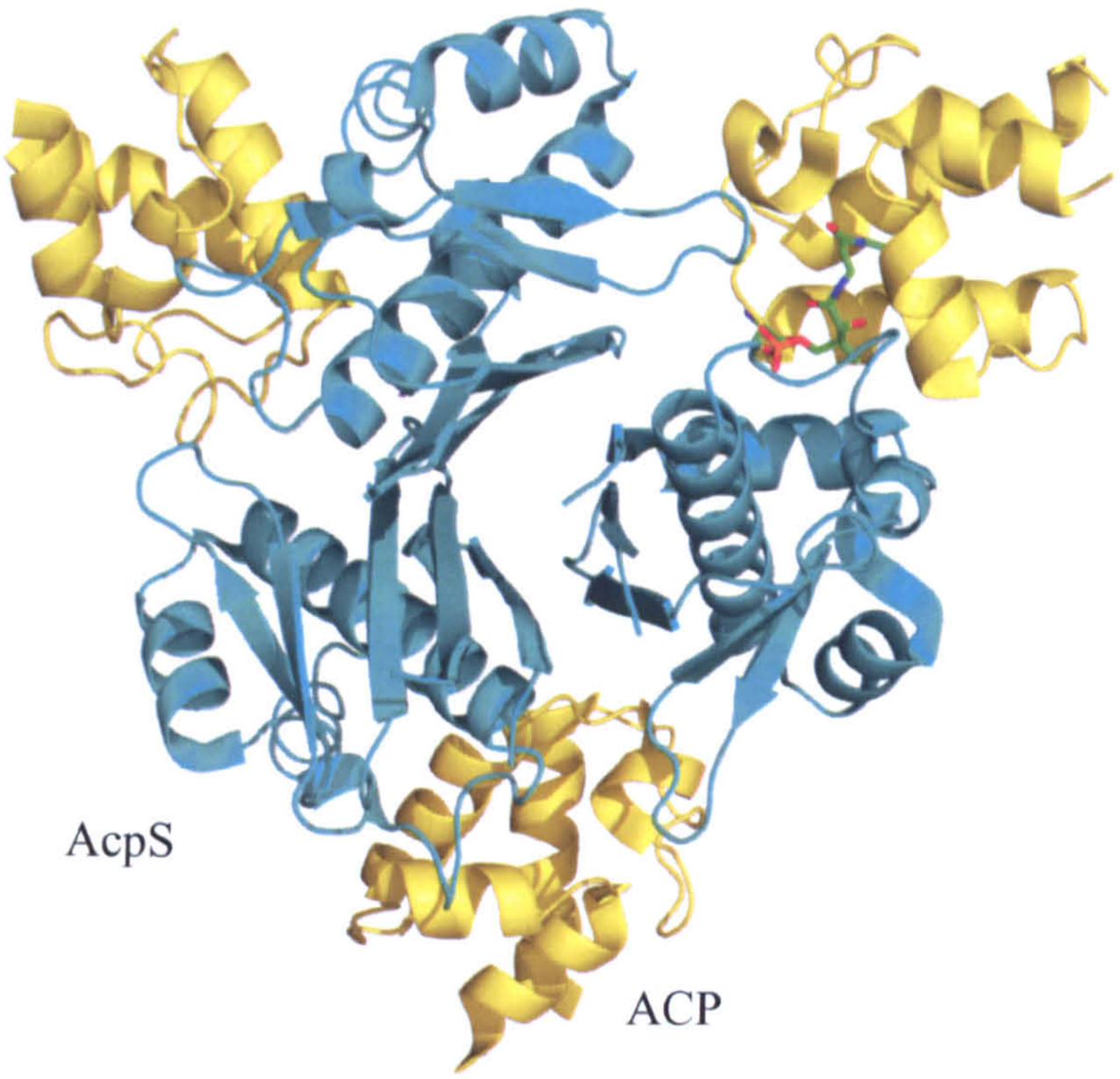


Figure 1.17 *B. subtilis* FAS AcpS:ACP complex. Cartoon representation of the *B. subtilis* AcpS:ACP complex (PDB code: 1f80). ACP is shown in yellow, AcpS in cyan, and the phosphopantetheine chain in green.

The structure reveals a novel trimeric arrangement of molecules, resulting in three active sites. The cofactor CoA lies at the interface between two monomers. ACP interacts with the AcpS via salt bridges formed by arginines in helix α1 of AcpS and glutamate and aspartate residues on helix α2 of ACP (Figure 1.18). It was proposed that a water molecule in the AcpS active site, named "activating water", can be deprotonated by Asp35 of ACP or by a solvent water and then deprotonate the hydroxyl group of ACP Ser36, the site for the attachment of the phosphopantetheine chain [86]. In addition, the structure of the *Streptococcus pneumoniae* AcpS has been solved in its apo form and in complex with the 3'-5'-ADP [87], and the structure of the human AcpS complexed with CoA and in ternary complex with CoA and human type I ACP has been recently determined [66] (Figure 1.19). *S. pneumoniae* AcpS is very similar to the *B. subtilis* phosphopantetheinyl transferase, with a trimeric fold and a cofactor binding site located at the interface between two monomers [87]. The human AcpS belongs to the type II class. It consists of two identical domains connected by a short linker

region, in a fold that is similar to the Sfp one, with the exception of the N- and C-terminal region that are unique to the human enzyme. The type I ACP domain was also solved in this study, but the ACP serine was mutated to an alanine to facilitate crystallization. The interaction between the two proteins occur on a large interface and it is predominantly hydrophobic [66]. This recognition is different from the one proposed for the *B. subtilis* complex, where negatively charged residues are proposed to mediate AcpS binding [86], although the positioning of the human ACP is similar to that for the *B. subtilis* fatty acid ACP. A comparison between human and *E. coli* ACPs revealed a considerably lower overall negative charge in the human ACP. It was proposed that, after magnesium and CoA binding, ACP completes the complex. A proton is then abstracted from the serine not by a water molecule, but by a glutamate residue from AcpS [66], in contrast to what was proposed for the *B. subtilis* complex.

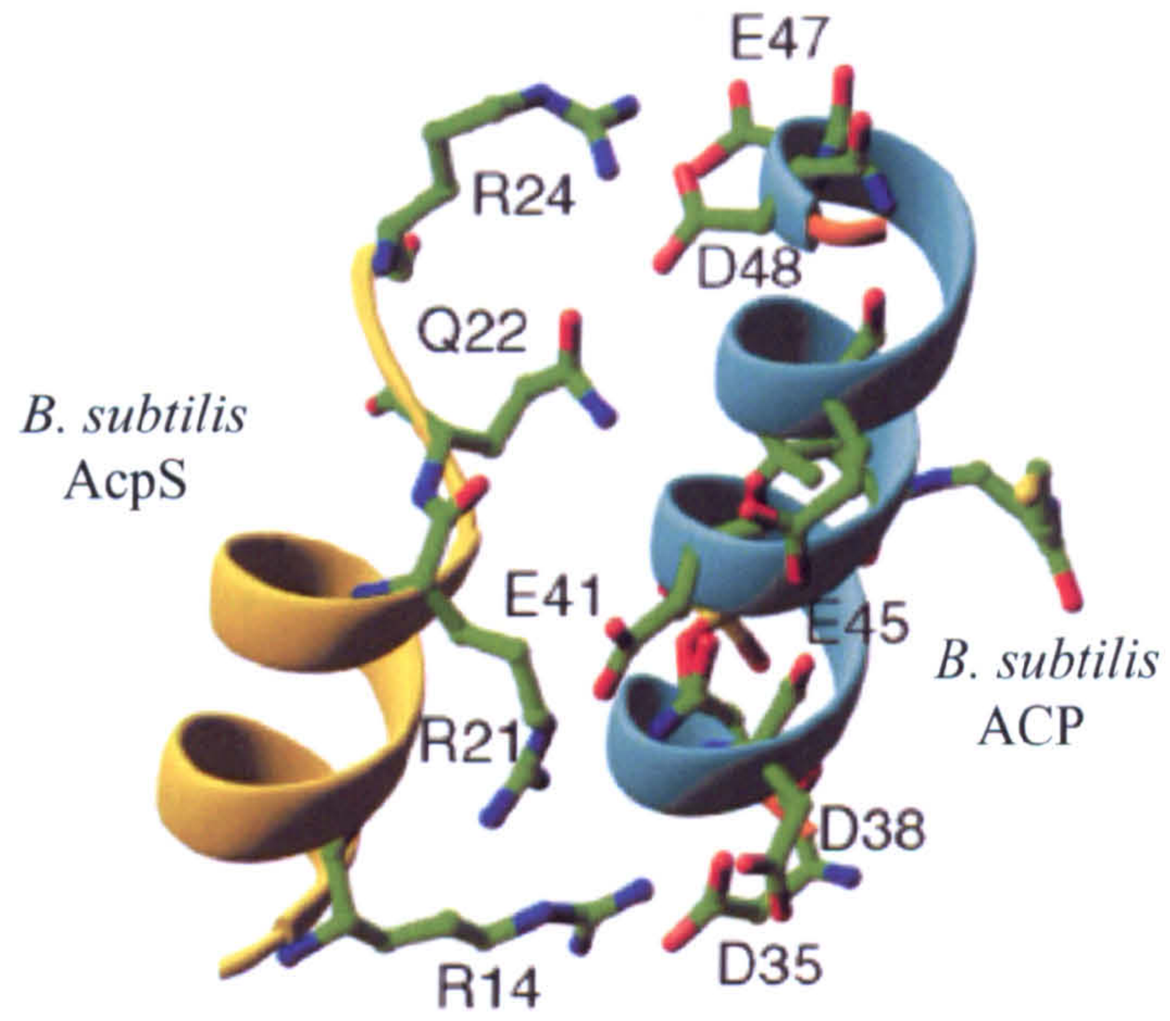


Figure 1.18 Interaction between *B. subtilis* AcpS and ACP. Conserved arginines in the AcpS make salt bridges with negative ACP residues (Figure adapted from [86]).

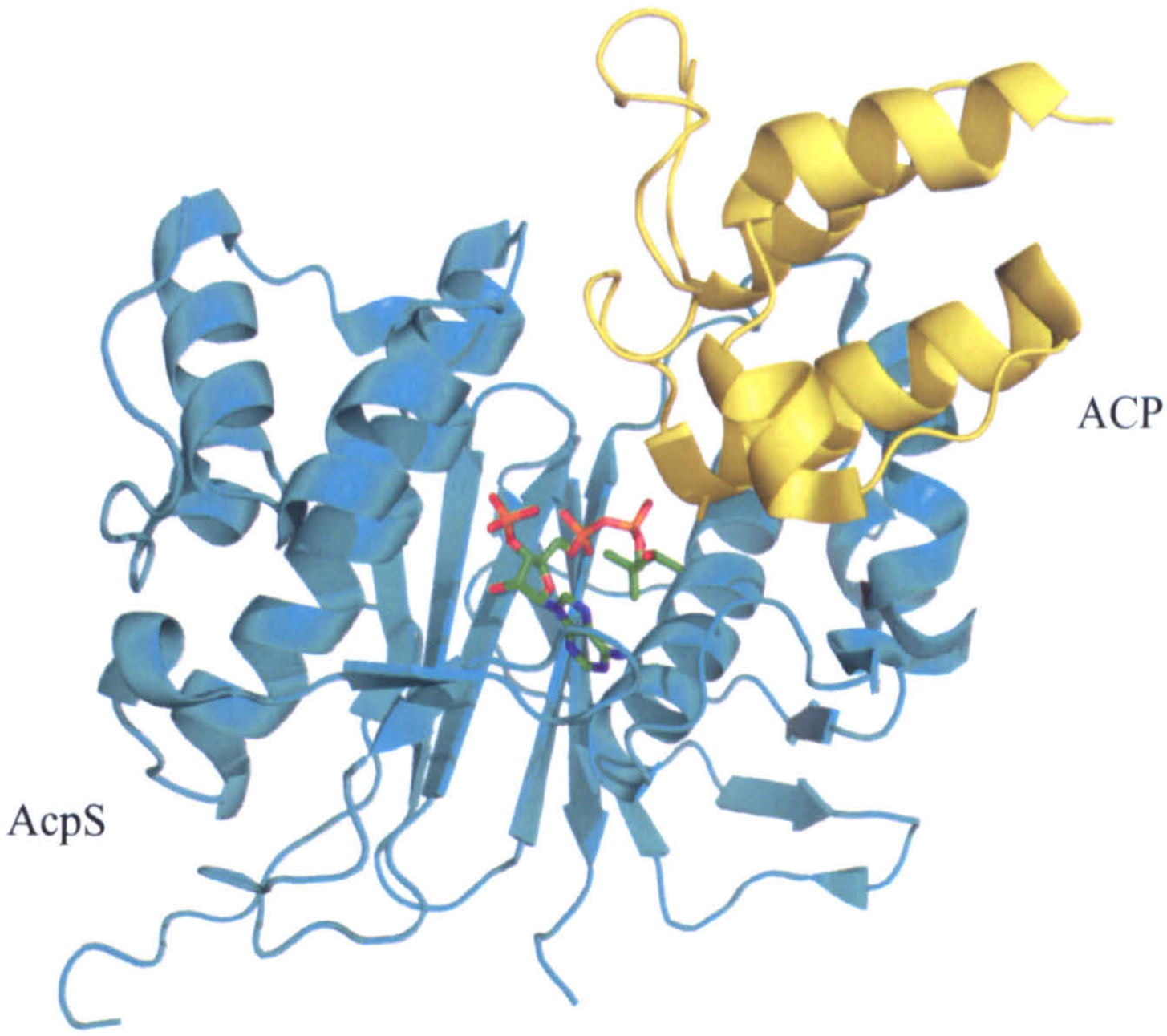


Figure 1.19 Human AcpS:ACP complex. Cartoon representation of the human AcpS in complex with type I fatty acid ACP domain and CoA (PDB code: 2cg5). AcpS in shown in cyan, ACP in yellow, and CoA in green.

1.4.3.3.1 *Streptomyces coelicolor* AcpS

The gram positive *Streptomyces coelicolor* produces the antibiotics actinorhodin and methylenomycin A [88], the immunosuppressive undecylprodigiosin [89], the non-ribosomal peptide antibiotic CDA [90-92] and also the usual range of bacterial fatty acids. All of these compounds require holo-ACPs for their biosynthesis. The gene cluster encoding for undecylprodigiosin synthesis appears to contain a phosphopantetheinyl transferase gene [89], but the other biosynthetic gene clusters do not, suggesting that *S. coelicolor* contains a multifunctional protein capable of activating a wide range of acyl proteins. *S. coelicolor* AcpS was identified, expressed and characterized by Cox and colleagues [74]. It was shown that this AcpS can transfer the phosphopantetheine chain to the fatty acid ACP as well as to the type I and II polyketide ACPs [74], a property that is quite unusual. In addition, it was proved that this protein can transfer modified CoAs to its ACP partner. *S. coelicolor* AcpS is therefore "promiscuous", in term of both substrate recognition and cofactor transfer. The *S. coelicolor* AcpS, in an activation reaction with apo-

ACP and modified CoAs, can be used for the production of acylated ACPs, which can be utilised for further investigation of the PKS.

1.4.3.4 Ketosynthase α and ketosynthase β

The ketosynthase α (KS_α) and ketosynthase β (KS_β) form a heterodimer in solution [93], and show high sequence similarity (Figure 1.20). KS_β is lacking an active site serine, which is essential for the activity. KS_α catalyzes the Claisen condensation, while the role of the KS_β subunit is not very clear and became the subject of scientific investigations. The important cysteine in the active site of the α subunit is mutated to a glutamine in the KS_β. Despite this, KS_β is not catalytically silent. This glutamine seems to be involved in loading malonyl-CoA and its decarboxylation to acetate (Figure 1.21).

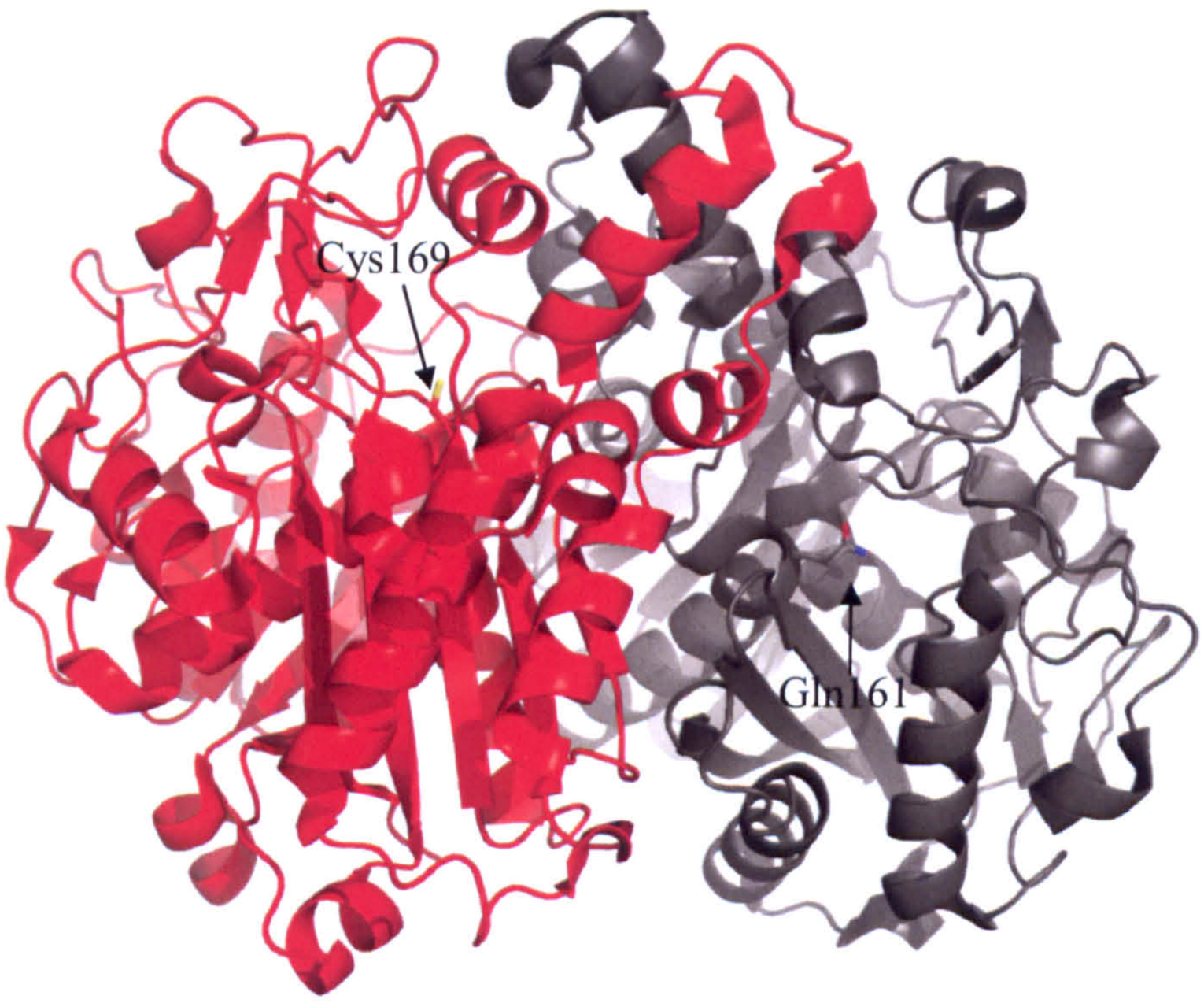


Figure 1.20 Type II actinorhodin KS_α/ KS_β. Cartoon representation of the actinorhodin ketosynthase/chain length factor. KS_α is coloured in red, KS_β in grey. The serine and the glutamine in both active sites are shown as sticks.

In addition, mutagenesis studies on the β-ketoacyl synthase unit of an animal fatty acid synthase, where the cysteine was replaced with a glutamine, showed that the rate of malonyl decarboxylation was increased by two orders of magnitude [94]. Consequently, it was proposed that the β subunit acts as a chain initiation factor that generates acetyl ACP, and for this reason it was also called chain initiating factor (CIF) [54]. In addition, KS_β is supposed to be the primary determinant of the right carbon length, and for this reason it is also known as chain length factor (CLF) [95, 96]. From the analysis of the actinorhodin KS-CLF crystal structure [93], it was proposed that a protein cleft keeps the growing chain extended. Several amino acids were identified that serves as gatekeepers in the polyketide tunnel. In addition, the first ring cyclization of the polyketide occurs within the KS-CLF tunnel, and it was also proposed that CLF is a function of the KS_α domain [93].

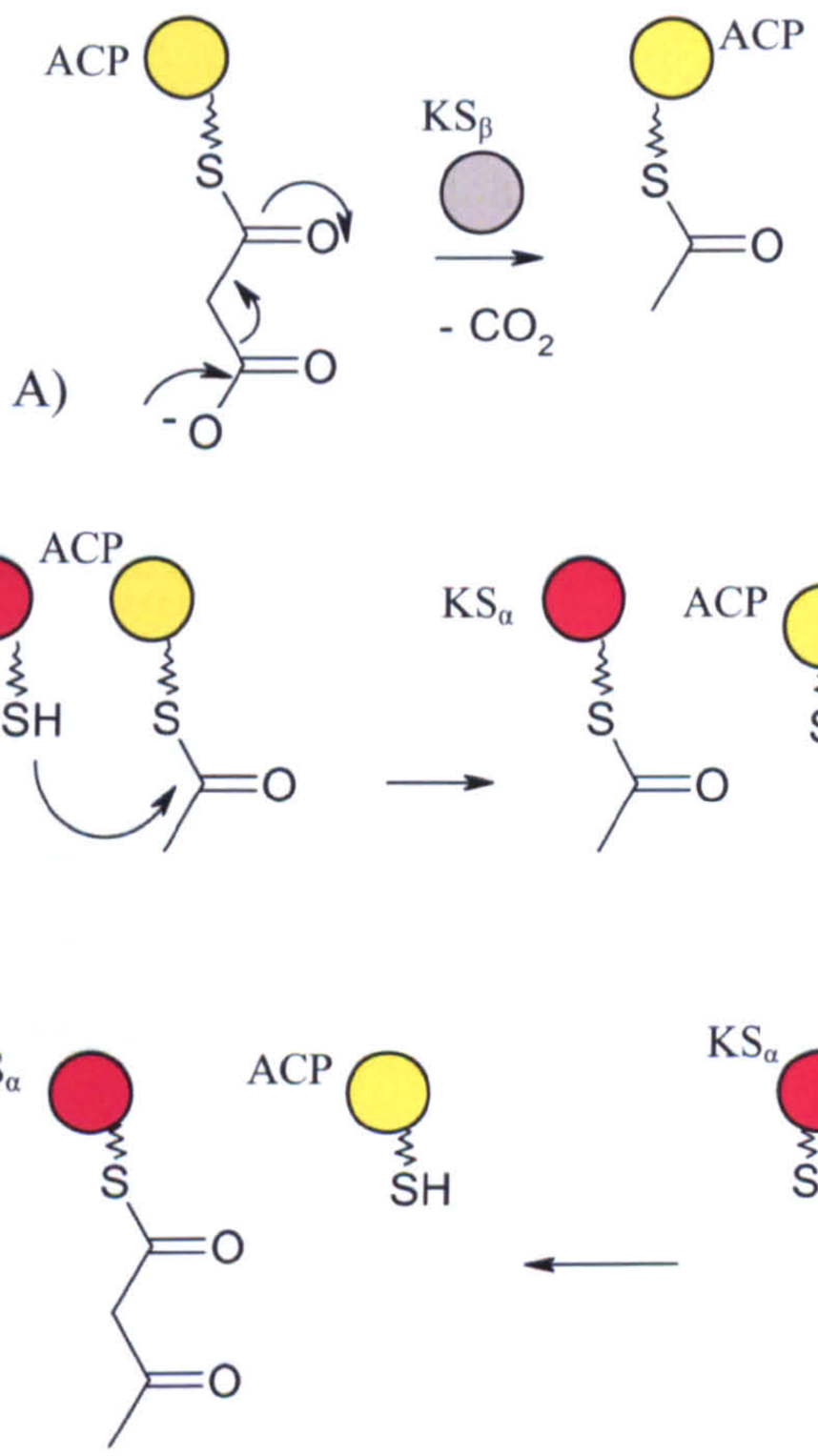


Figure 1.21 Reaction catalyzed by the KSα and KS_β complex. A) Acetyl starter unit is formed by decarboxylation of malonyl-CoA catalyzed by KS_β. B) The starter unit is transferred to the KS_α, while the ACP is loaded with the malonyl extender unit. Cycles of condensation result in a growing chain that is transferred from ACP to KS_α.

1.4.3.5 Ketoreductases

Ketoreductases (KRs) catalyze the stereospecific hydrogen transfer from NAD(P)H onto a keto group. In most cases, they are an integral part of bacterial aromatic PKSs, where they can be the first enzyme to modify the nascent polyketide chain prior to cyclization [97].

1.4.3.5.1 Actinorhodin ketoreductase

The minimal system responsible for the synthesis of actinorhodin produces a 16 carbon polyketide chain, and can also partially control the first cyclization, though aberrant products with the wrong cyclization can be observed [53, 98]. The actinorhodin ketoreductase (KR) is the first enzyme that modifies the nascent polyketide, reducing the carbonyl group at C-9. The act minimal system produces two compounds, SEK4 with the correct first cyclization (C7-C12), but with the other incorrect, and SEK4b - the intermediate preceding mutactin - with the alternate (C10-15) cyclization [53]. The addition of the KR to the minimal system, *in vivo* and *in vitro*, predominantly yields mutactin, which is correctly cyclised and reduces the presence of incorrectly cyclised intermediates, SEK 4b and SEK4 [99, 100] (Figure 1.22). This is good evidence that KR can help the minimal system in catalysing the correct cyclization C7-C12.

Moreover, in presence of KR there is an increase in the overall polyketide production in comparison with the minimal system, and in KR-deficient mutants there is an increase in misfolded polyketide [101, 102]. Both pieces of evidence could suggest that KR may stabilize the complex or produce more efficient intermediates.

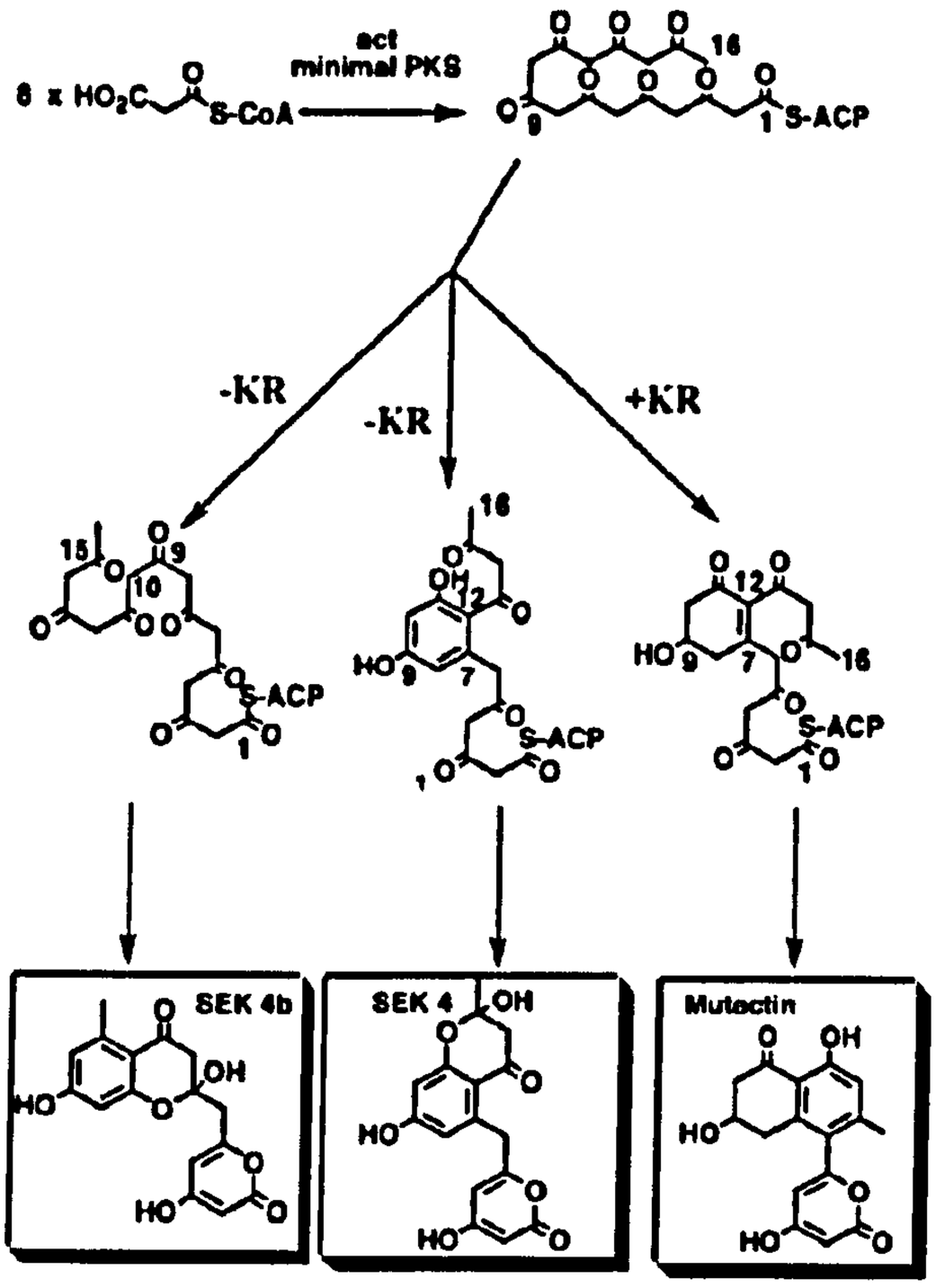


Figure 1.22 Putative biosynthetic scheme for the early steps of actinorhodin biosynthesis. A mixture of SEK4 and SEK4b is produced by the minimal system, and the addition of the KR leads to production of mutactin (Figure adapted from [103]).

Recently, the first structure of a type II ketoreductase, the actinorhodin KR from *Streptomyces coelicolor*, was solved at a resolution of 2.5 Å [103]. Act KR belongs to the short-chain dehydrogenase/reductase evolutionary family (SDR) of NADP(H)-dependent enzymes [24, 104-106]. The most closely related sequences are those of fatty acid ketoreductases, which reduce the beta-keto carbonyl of the beta-ketoacyl-ACP in the first reductive stage of fatty acid synthesis [107-110]. In the crystal structure (Figure 1.23), the enzyme is a tetramer made up of a dimer of dimers, and each crystallographic unit contains 2 subunits, A and B, with two different conformations. Both subunits are arranged in a Rossmann dinucleotide binding fold, each containing a 7 stranded β-sheet and 8 α-helices, and bound to the unreduced cofactor, $NADP^+$ [103].

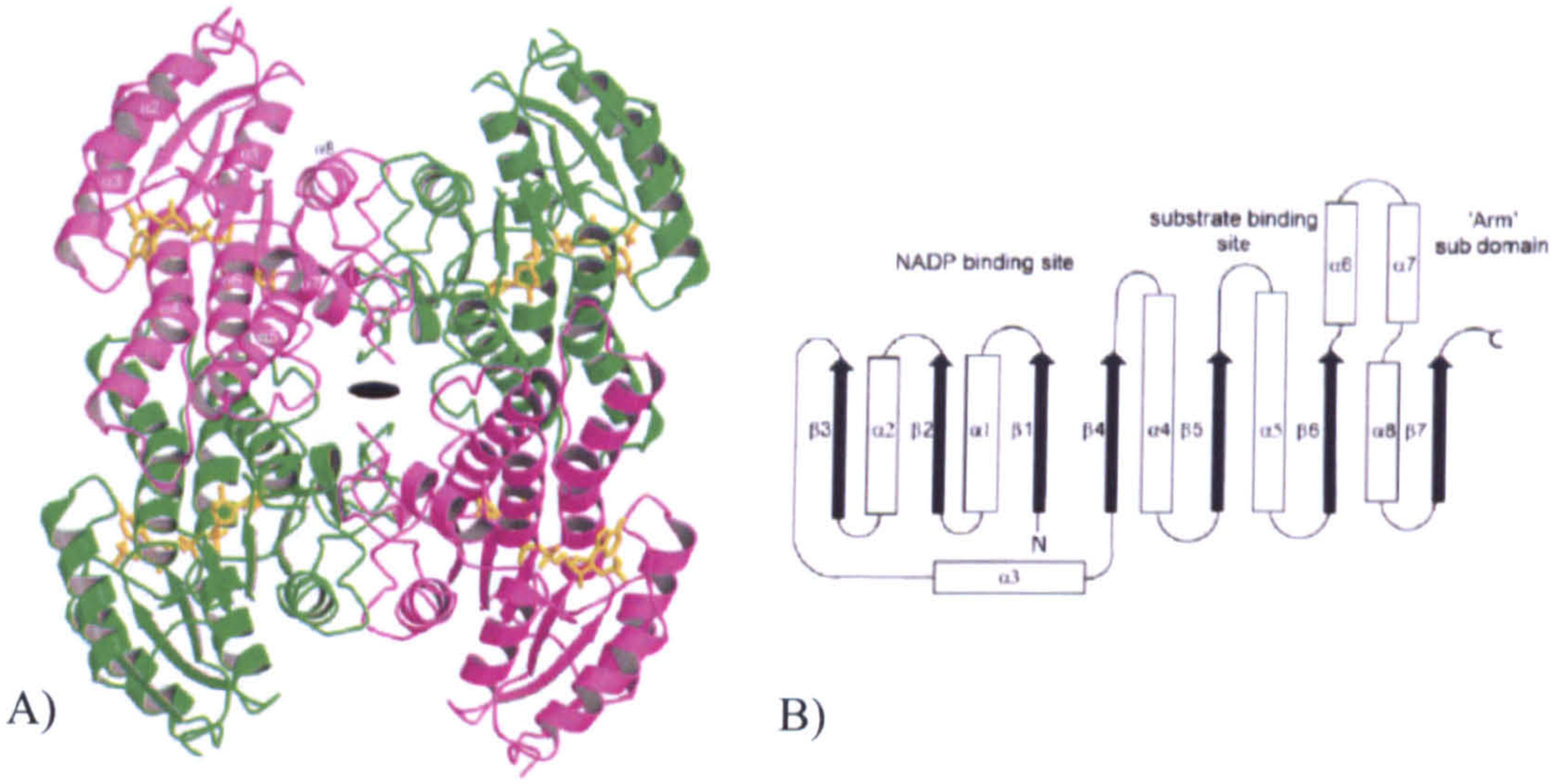


Figure 1.23 Actinorhodin ketoreductase crystal structure. A) Tetrameric arrangement of KR, with the A subunit in magenta and the B subunit in green. B) Topology diagram (Figure adapted from [103]).

The act KR was crystallized in the presence of 4 M sodium formate, and formate molecules are found in the active site of the open subunit A, while in the closed subunit B they are not present. These formate molecules could indicate favourable binding sites for the carbonyl oxygens of the nascent polyketide.

The active site is composed of a conserved triad of residues, which consists of serine, tyrosine and lysine, and it is characteristic of the whole SDR family [111], while the substrate binding site is located above the $NADP^+$ and toward the C-terminal end of the subunit, near helices α6 and α7. In SDR enzymes, the reduction catalytic mechanism is proposed to occur via hydride transfer from the NADPH followed by proton donation from a nearby tyrosine hydroxyl group. This proton transfer occurs from the tyrosine of the enzyme to the carbonyl oxygen of the substrate. The conserved serine is thought to help stabilize the bound substrate through hydrogen bonds. The amino group of the lysine can provide hydrogen bond interaction with the nicotinamide ribosyl moiety of the cofactor and can also lower the pK_a of the proton on the catalytic tyrosine through ionic interactions [108, 111]. By superimposing the two subunits, it appears clear that there is a structural difference in the conformation of helices α6 and α7, that, compared to subunit B, appear to be rotated 19° in subunit A, and resulting in a more open active site

conformation for the subunit A, while in the subunit B the closed position of these helices seems to restrict the access to the active site (Figure 1.24).

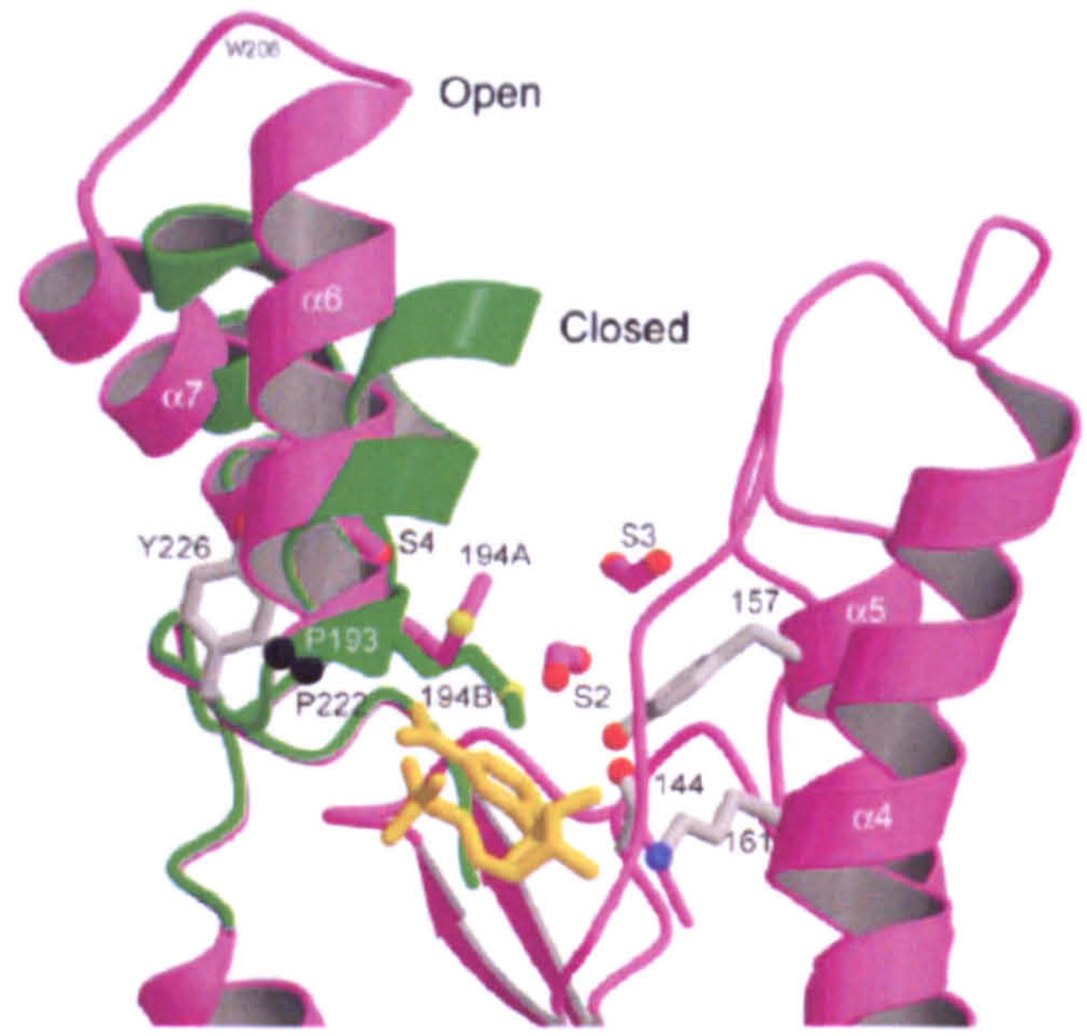


Figure 1.24 KR open and closed forms. Superimposition between the open A subunit (magenta) and the closed B subunit (green) (Figure adapted from [103]).

It was suggested that the active site is opened by the binding of the cofactor [110], but the fact that both subunits are bound to $NADP^+$ in the act KR seems to contradict this hypothesis. It appears instead that a crystal contact between the A subunit and a symmetry related B' subunit leads to the opening, in particular helices αa and αb of the B' subunit impinge upon both the cofactor binding region and the active site of subunit A [103]. So, protein-protein interaction seems to be the mechanism for the opening of the subunit A. In particular, there are features of the interacting part of subunit B that are reminiscent of ACP. A model is therefore proposed for the interaction between KR and ACP. The cofactor should bind to each subunit, stabilising the formation of the tetramer. In this state, the catalytic triad is placed in the right position and the substrate binding site is closed with water excluded. Then a negative charged patch on helix α2 of ACP can recognize the positive patch present on KR surface, inducing the conformational change that allows the insertion of the substrate in the active site. Inspection of this docking model revealed that negative glutamate ACP residues were located in a positive patch formed by arginines 38, 65 and 93. On the contrary, for the *E. coli* KR FabG,

arginines 129 and 172 were proposed to be essential for ACP binding. Mutations of these two residues abrogate the fatty acid biosynthesis [25]. These two residues are not conserved among PKS KRs, and their closest counterparts are far away from the active site [103].

1.4.3.6 Cyclases and aromatases

Highly reactive poly-β-keto intermediate can undergo spontaneous aldol reaction or directed cyclization. Cyclases (CYCs) function in a chaperone-like way, directing the intermediate into particular channels were aldol chemistry is suppressed. Unfortunately, the instability of their substrates complicates their study. So, their function can only be understood through inactivation experiments and analysis of the products. Anyway, it seems that cyclization occurs before the dehydrogenation carried out by aromatases [29]. Cyclases can exist as didomain with internally duplicated motifs [52, 112], or as monodomains [113]. These domains can be fused with other functionalities, for example CYC/KR or CYC/MT. An example of cyclise is given by the tetracenomycin F2 cyclase from *S. glaucesens*, TcmI. The crystal structure revealed an arrangement similar to the one found in other polyketide monoxygenases, with a large cavity where the substrate should bind [114].

In addition to cyclases, aromatases help the cyclization process. These enzymes dehydrate cyclic alcohols to yield aromatic rings. A particular example is given by LanV involved in landomycin biosynthesis. The protein appears to be bifunctional, catalysing the ketoreduction step at position C-6 followed by the dehydration reaction and thus the aromatization of the ring A of angucycline [115]. The crystal structure of the tetracenomycin (tcm) ARO/CYC was solved recently [116]. This enzyme contains a highly conserved interior pocket. The size, shape and composition of the pocket are proposed to be important to orient and specifically fold the polyketide chain for C9-C14 first-ring and C7-C16 second-ring cyclizations. It was also proposed that the regiospecific cyclizations of the first two rings and subsequent aromatizations take place in the interior pocket.

1.4.3.7 Tailoring enzymes

In addition to the enzymes previously described, polyketide synthesis is also characterized by the presence of tailoring enzymes that extend the structural diversity of the products. Methyltransferases usually use the S-adenosyl-L-methionine (SAM) cofactor to transfer its activated methyl group to nitrogen, carbon or oxygen atoms. The degree of amino acid similarity among these proteins is quite low, although they often share a common chain-fold consisting of a central β-sheet with surrounding α-helices, with a SAM binding site. Oxygenases promote the incorporation of oxygen into a substrate. They are divided into mono or dioxygenases, depending on the number of oxygens transferred. Several types of oxygenases are known to act in the post-PKS II modification, such as the cytochrome P-450 monooxygenases, flavin-dependent mono- and dioxygenases, and anthrone oxygenases. Finally, glycotransferases can catalyse the formation and attachment of sugars to bacterial aromatic polyketides [29].

1.4.3.8 Daunorubicin/doxorubicin synthesis

Daunorubicin belongs to the anthracycline family and it is used for the treatment of cancer. It was initially isolated form *Streptomyces peucetius* and it is most commonly used to treat specific types of leukemia. The hydroxylation of carbon C14 yields doxorubicin, a drug widely used in cancer therapy. Daunorubicin is more abundant as natural product, and can be produced by a wide range of *Streptomyces* strains, while doxorubicin can be produced only in special environmental conditions or by introduction of genetical modifications [117, 118]. The cytochrome P450 oxydase, DoxA, catalyses the oxidation of daunorubicin to doxorubicin [119, 120]. The anthracycline skeleton of doxorubicin is produced by a type II polyketide synthase. A 21-carbon decaketide chain is produced by condensation of a single propionyl-CoA and 9 malonyl-CoA. After the elongation steps are concluded, successive modifications produce a tetracyclic anthracycline aglycone [121]. The daunosamine amino sugar is attached to the aglycone and other modifications yield daunorubicin and finally doxorubicin [122]. *Dps* genes are responsible for the linear chain formation and the first cyclizations. The *dnr* and *dnm* gene clusters are responsible for the remaining modifications and for the amino sugar synthesis respectively. After the reactions catalyzed by the minimal

system DpsA (KS_α), DpsB (KS_β) and DpsG (ACP), with the additional help of DpsC and DpsD, the ketoreductase DpsE is the first enzyme that catalyzes reduction in position C9. It was suggested that DpsC, a FabH-like ketosynthase, is responsible for the selective choice of propionyl as a starter unit and catalyzes the first condensation of malonyl-CoA with propionyl-CoA [123, 124], and then DpsD, an acyl-transferase, loads the diketide onto the minimal complex [118, 125]. After genetic modification experiments, it was also revealed that DpsE requires substrates with a chain length of more than 16 carbon atoms [126]. The catalysis of daunorubicin is summarized in Figure 1.25.

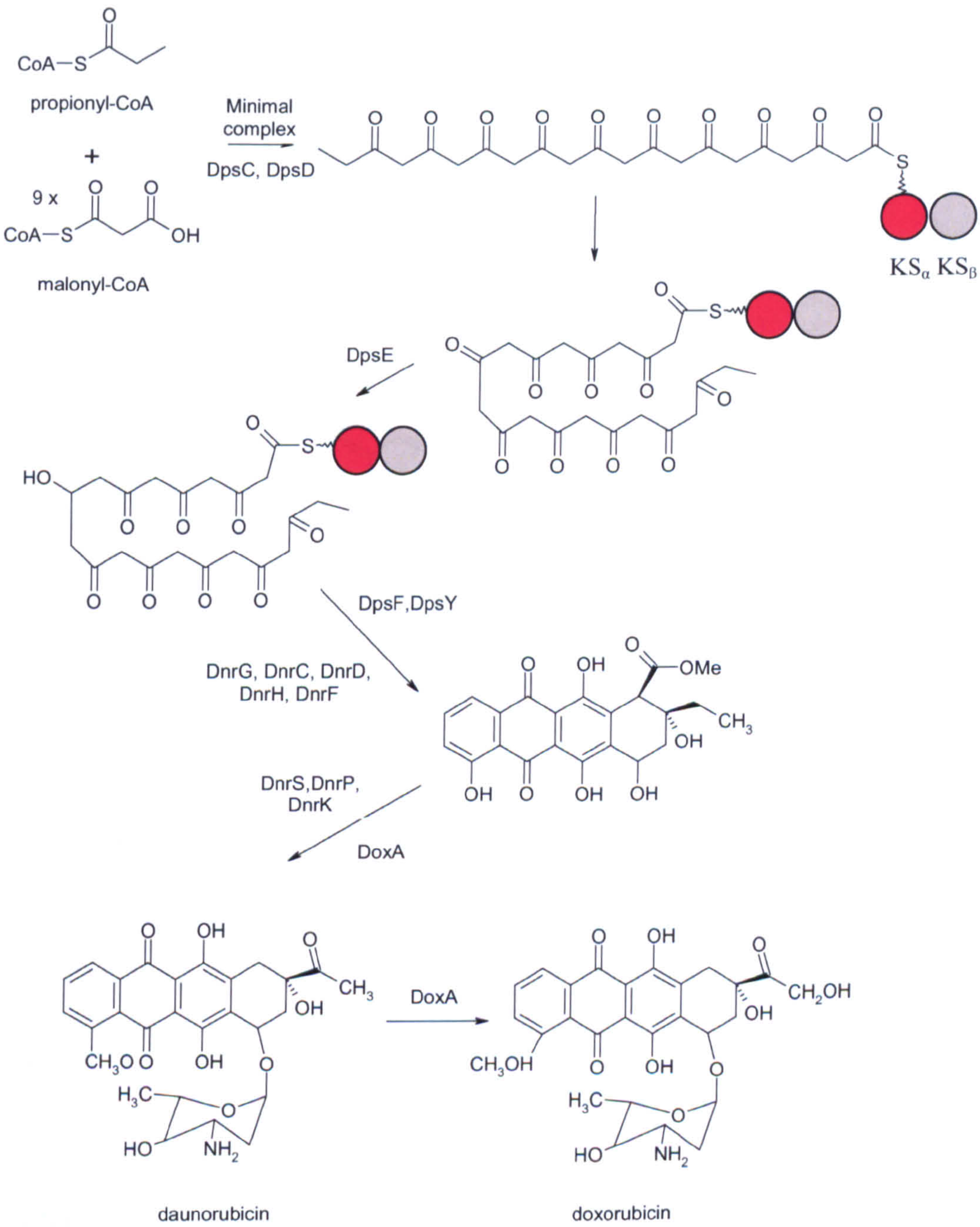


Figure 1.25 Daunorubicin/doxorubicin polyketide synthase.

1.4.4 Type III PKSs

Type III PKSs are found in bacteria and higher plants. They are typically multifunctional enzymes that catalyse the synthesis of important intermediates for the production of floral pigments, phytoalexins and mediators of plant fertility, and melanin and other napthoquinone ring based metabolites in various filamentous bacteria [30, 31, 127, 128]. Examples are the chalcone and stilbene polyketide synthases (Figure 1.26). They are giant condensing enzymes that catalyze the condensation of p-coumaroyl-CoA and malonyl-CoA to a common tetraketide intermediate, followed by different cyclization reactions that produce naringenin chalcone and resveratrol respectively [129-131].

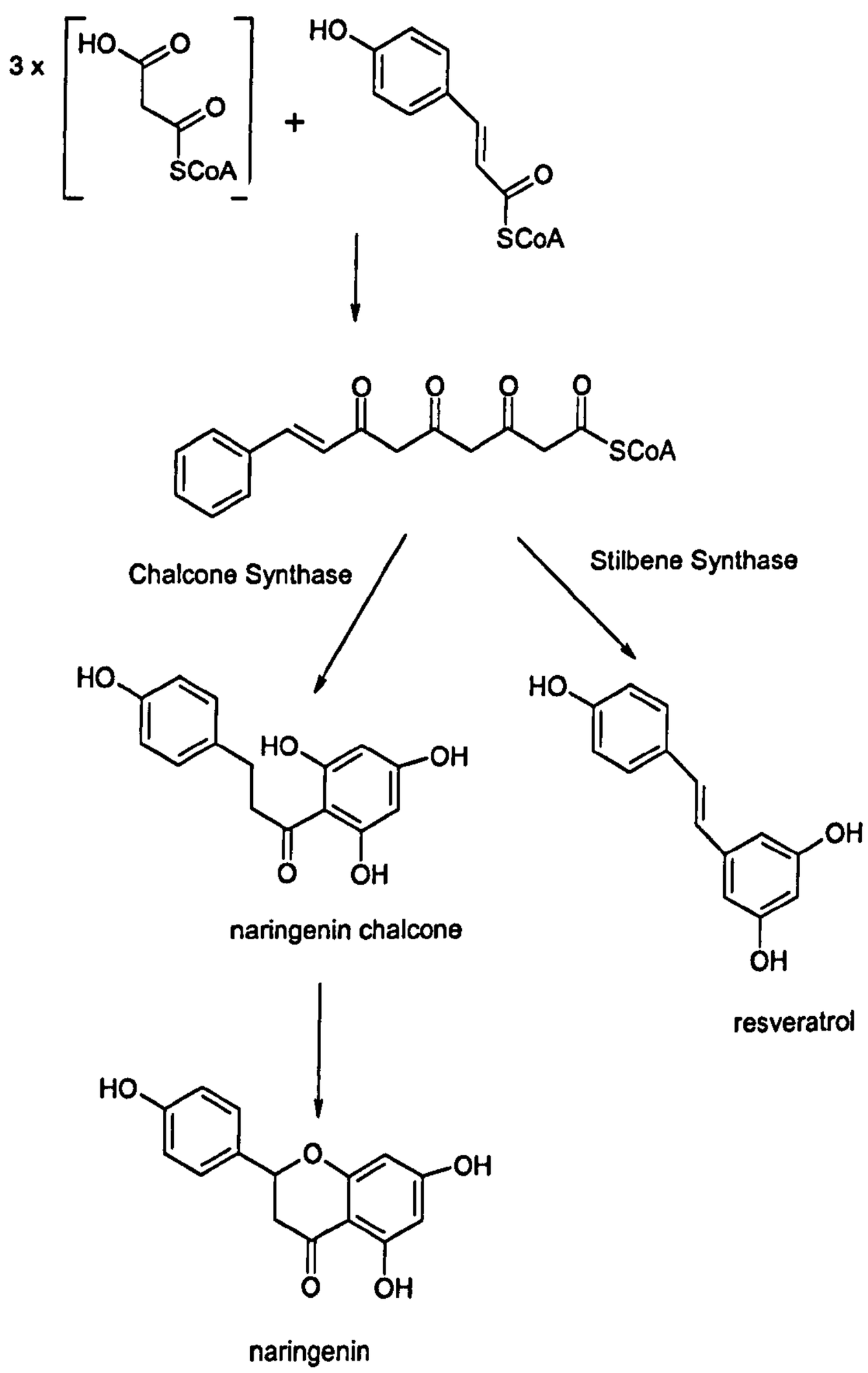


Figure 1.26 Reactions catalyzed by type III chalcone and stilbene polyketide synthases.

These PKSs utilize CoA thioesters directly, without using the phosphopantetheine arm of ACP, moreover, they catalyze all the decarboxylation, condensation, cyclization and aromatization reactions in a single active site. The *Medicago sative* chalcone synthase crystal structure was solved in 1999 [127]. The protein is related to the *E. coli* KASIII and the thiolase super family. In the active site, the conserved residues Cys164, Phe215, His303 and Asn336 were identified. It was proposed that the thiolate of the cysteine acts as a nucleophile to load the acyl intermediate, the asparagine and histidine carry out the decarboxylation of malonyl-CoA, and the phenylalanine orients the substrates during the elongation cycles [32] (Figure 1.27).

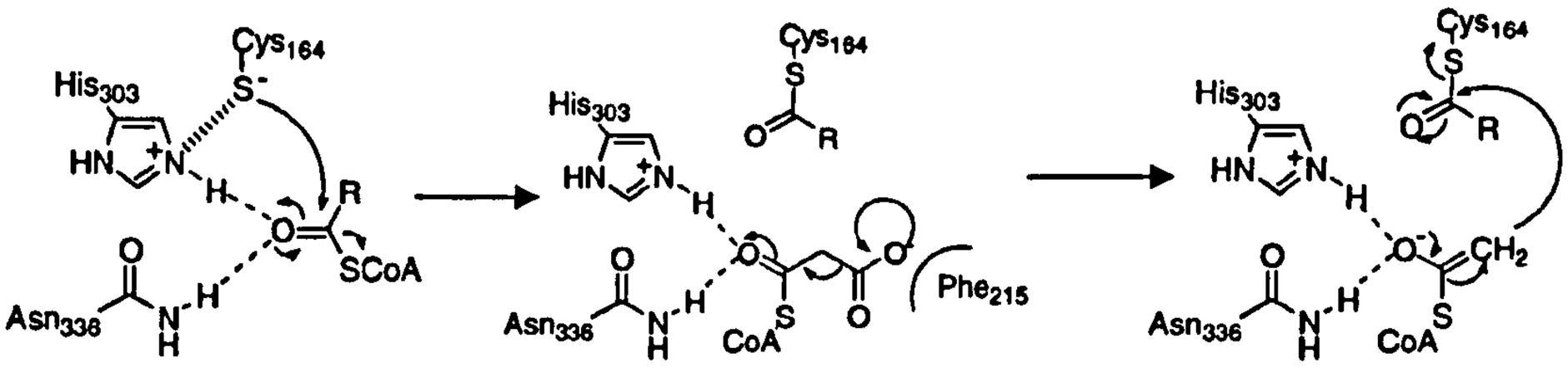


Figure 1.27 Proposed mechanism for the chalcone synthase. R is the coumaroyl moiety in the first reaction cycle, coumaroyl-acetyl in the second, and coumaroyl-diacetyl in the third. Figure adapted from [3].

1.5 Objectives and aims

The aim of the work carried out during the PhD was to increase the understanding of type II polyketide synthesis. The main areas of the project have included:

1. analysis of the crystal structure of the *Streptomyces coelicolor* acyl carrier protein synthase (AcpS), and its comparison with other type I and II AcpSs;
2. investigation of the substrate promiscuity of the AcpS, in particular analysing the crystal structure of the enzyme in complex with acetyl-CoA, and the substrate binding sites of other type I and II AcpSs;
3. proposals about the *S. coelicolor* AcpS:ACP recognition and interaction, and about the phosphopantetheinylation mechanism;
4. generation and analysis of the crystal structures of H110A and D111A AcpS mutants, looking at the role and at the effects that these residues could have in catalysis.

MATERIALS AND METHODS

Chapter 2

Methods

2.1 Expression vectors

A plasmid vector is an extra chromosomal DNA element that can replicate independently of the chromosomal DNA. It is commonly present in prokaryotes and usually it is a circular double stranded DNA. A gene that encodes the protein of interest, can be cloned in a well determined position without altering the functionality of the rest of the plasmid. The majority of expression vectors are produced by natural plasmids, and they usually have the following features for a correct replication and protein expression:

- an origin of replication (ori) for replication of the plasmid in the host;
- one or more selection markers, such as antibiotic resistance genes, in order to select the bacteria that contain the plasmid;
- a Multiple Cloning Site (MCS), characterized by unique cleavage sites, in order to clone the gene in a defined position;
- a promoter, usually from a bacteriophage, capable of being recognized by a specific RNA polymerase, and controlled by a specific system for a controlled protein expression;
- a repressor gene sequence, that can bind to the operator site before the promotor, preventing the RNA polymerase from binding and starting the gene transcription. A classical example is the *lacI* repressor, which binds to the O_{lac} operator. The repression can be switched off by the inducer isopropyl-β-D-thiogalactoside (IPTG);
- a Ribosome Binding Site (RBS), situated before the first ATG codon;
- an optional gene (tag) that allows the expression of an amino acid sequence (His_6-*tag*) or a polypeptide (GFP), on the N- or C- terminus of the protein, in order to help the subsequent purification.

2.1.1 pET vectors

The pET vectors (Novagen) are among the most common systems for the cloning and the expression of proteins in *E. coli*. Target genes are cloned under the control

of the strong bacteriophage T7 promoter and the expression is induced by a T7 RNA polymerase provided by the host cell. In particular, these plasmids carry a f1 origin of replication; the T7 promoter and terminator; a *lacI* gene coding for the *lac* repressor protein; a *lac* operator that can block transcription; a Multiple Cloning Site; the ampicillin or kanamycin resistance gene. In addition, with the exception of pET-3d, they carry a His_6-*tag* and a thrombin site coding region. Vector maps are presented in Appendix 2.

2.2 *E. coli* BL21 star (DE3)

The *Escherichia coli* BL21 star (DE3) strain is a very common host for the expression of proteins that are cloned into pET vectors. These bacteria contain the λDE3 lysogen phage, which carries the gene for the T7 RNA polymerase, the latter controlled by the *lac*UV5 promoter. IPTG inhibits the *lacI* repressor and is required to induce the expression of the T7 RNA polymerase (Figure 2.1). Furthermore, the *lon* protease and the outer membrane protease OmpT are missing, therefore reducing the possible protein degradation. A mutated *rne* gene (*rne*131) also encodes for a truncate RNase E with a reduced capability of mRNA degradation.

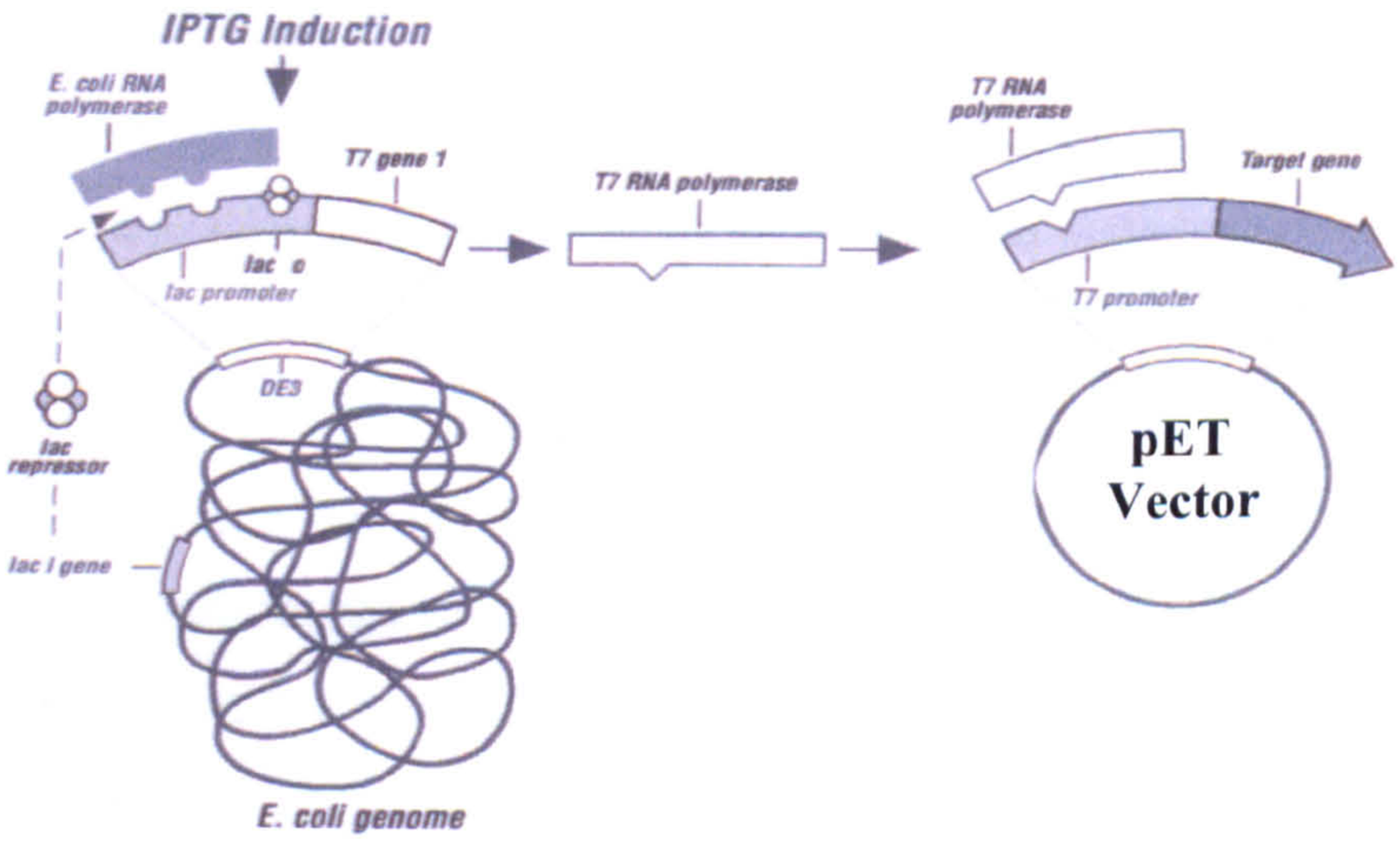


Figure 2.1 Expression of a target protein cloned into a pET vector. The lacI repressor precludes the expression of the T7 polymerase. Following the IPTG induction, the lacI repressor is inhibited by IPTG, the T7 RNA polymerase is expressed and so the target protein, controlled by the T7 promoter. (Figure adapted from Novagen manual).

2.3 PCR

The Polymerase chain reaction (PCR) is an *in vitro* enzymatic replication used to amplify a DNA template. The reaction is characterized by several thermal cycles of alternately heating and cooling, and the following reagents are necessary:

- the DNA template containing the target region to be amplified;
- two primers, single strand DNA fragments complementary to both sides (5' and 3') of the target DNA region. The optimal length for a primer is around 18-22 bp, with a GC content of 40-60% to avoid the formation of secondary structure, and they should end with a C or better G base, as this guarantees a strong binding at the 3' end;
- the DNA polymerase, like the Taq polymerase, is the extending enzyme working at a high temperature;
- dNTPs, the building blocks;
- buffer solution to ensure the right chemical environment for the polymerase;
- divalent cations, like magnesium or manganese ions, to stabilize dsDNA.

A PCR reaction involves 3 steps: denaturation, annealing and extension. The temperature and time employed for each cycle depends on the length of the template (usually 1 kb/minute), the particular polymerase used, the concentration of cations and dNTPs, and the melting temperature (T_m, the temperature at which one half of the DNA template dissociates to become single stranded) of the primers. During the denaturation step, usually at 95°C, the two strands of the DNA template are separated. Primers binds to each single strand DNA during the annealing step, at a temperature that is around 2-3°C below the T_m of primers. Finally, at the optimal temperature for the particular polymerase used (usually between 70-80°C), the new DNA strand complementary to the template is synthesized during the extension step in the 5'-3' direction. These 3 cycles are usually repeated 20-40 times.

2.4 Stratagene site-directed mutagenesis

Site-directed mutagenesis is a powerful and very precise technique that makes possible the analysis of the function of a particular amino acid residue. The technique allows the introduction of almost any mutation at any site. Several

protocols are available, but they require a single strand DNA as template. In contrast, the Stratagene QuickChange II site-directed mutagenesis allows site specific mutations in any dsDNA template. A dsDNA containing the target gene and the two primers, which enclose the mutated sequence, are used in a PCR reaction. The PCR is performed with the *PfuUltra* high-fidelity (HF) DNA polymerase. The extension of the primers produces a mutated plasmid carrying the desired mutation. The product is then treated with the *DpnI* endonuclease (target: 5'-Gm^{6}ATC-3'), which is specific for methylated and hemimethylated DNA, and cleaves the parental DNA. DNA isolated from *E. coli* is methylated by the Dam methylase and therefore subject to *DpnI* cleavage. The mutated plasmid is transformed in XL1-Blue supercompetent cells for an optimal replication of the vector.

2.5 Chromatography techniques

The first chromatography technique was invented in 1903 by the Russian botanist Tswett, who noticed the formation of coloured bands on a liquid absorption column while separating different plant pigments. The most important separation and purification techniques are based on chromatographic processes, where the target mixture, the mobile phase, is suspended as a gas (gas chromatography) or a liquid (liquid chromatography) and introduced into columns provided with solid particles, the stationary phase. The process of purification is based on the interaction between the two phases, delaying the elution of the analyte in a way that is strictly dependent on the property of the analyte itself. If the initial mixture is pictured like a narrow band, the different interactions with the stationary phase determine a different and characteristic migration speed for the target analyte, transforming the narrow band in a series of different bands composed of pure solutes.

Chromatography separations can be compared to organic extractions, although they are characterized more by absorption processes rather than partition phenomena that are common for small molecules. The retention process for each analyte is due to the equilibrium that the analyte establishes with the mobile and the stationary phase. This equilibrium is described by the equilibrium constant or partition coefficient K_{eq}, defined as the molar concentration of analyte in the stationary phase divided by the molar concentration of the analyte in the mobile phase.

Various chromatographic parameters must be considered during a purification, such as:

- the dead volume V_0, or the volume of mobile phase between the sample injection and the peak maximum of the solute that is not retained;
- the retention volume V_{Ri}, the volume of mobile phase at the analyte peak maximum;
- the base width W_i for each peak;
- the capacity factor $k_i = \frac{V_{Ri} - V_0}{V_0}$.

The separation between each solute depends on both the width and the distance of the peaks. The separation of large bands requires more time, so smaller peaks are preferred. This preference is described by the efficiency or number of theoretical plates $N = 16 \left(\frac{V_{Ri}}{W_i}\right)^2$. This value is obtained from the parameters that characterize each band, creating a relation between the band width and the time of retention ($V_{Ri} = t_{Ri}$). A column can be seen as a number of theoretical plates, or imaginary layers where the separation occurs.

All these parameters describe the properties of each band. The following parameters are used to describe the properties that relate two or more peaks:

- the selectivity factor $\alpha_{12} = \frac{k_2}{k_1}$ (always greater than 1);
- the resolution $R = \frac{V_{R2} - V_{R1}}{\frac{1}{2}(W_2 - W_1)}$.

Resolution offers a quantitative measure of the separation grade between two peaks. For R=1.5, the overlapping region is about 0.1%. If the efficiency is a measure of the peak enlargement and is determined by the property of the column (type of matrix and packing) used and by the rate of the mobile phase, the resolution is determined by the properties of both phases, and by the temperature and length of the stationary phase. Another important parameter is the height equivalent to a theoretical plate HETL = L/N, where L is the length of the column. The best chromatographies are obtained with low HETL values.

Finally, the best model used to describe the creation of a band is given by the van Deemter equation. The band width reflects the distribution of the molecule that has been eluted in the system. The model describes the variation of HETL in relation to

the linear average rate of the mobile phase v: $H = A + \frac{B}{v} + Cv$ (Figure 2.2). A, B and C are constants associated to physical properties of the mobile and stationary phases. The equation describes the enlargement of the chromatographic bands. Contributions are given by the longitudinal diffusion (B), by the Eddy diffusion (A), related to the packing of the column, and the resistance to the mass transfer (C), to which contribute both the mobile phase rate and the distribution of the dimension of the pores in the stationary phase.

The different chromatographic techniques are classified on the base of the properties of the two phases and of the nature the interaction between solute and stationary phase.

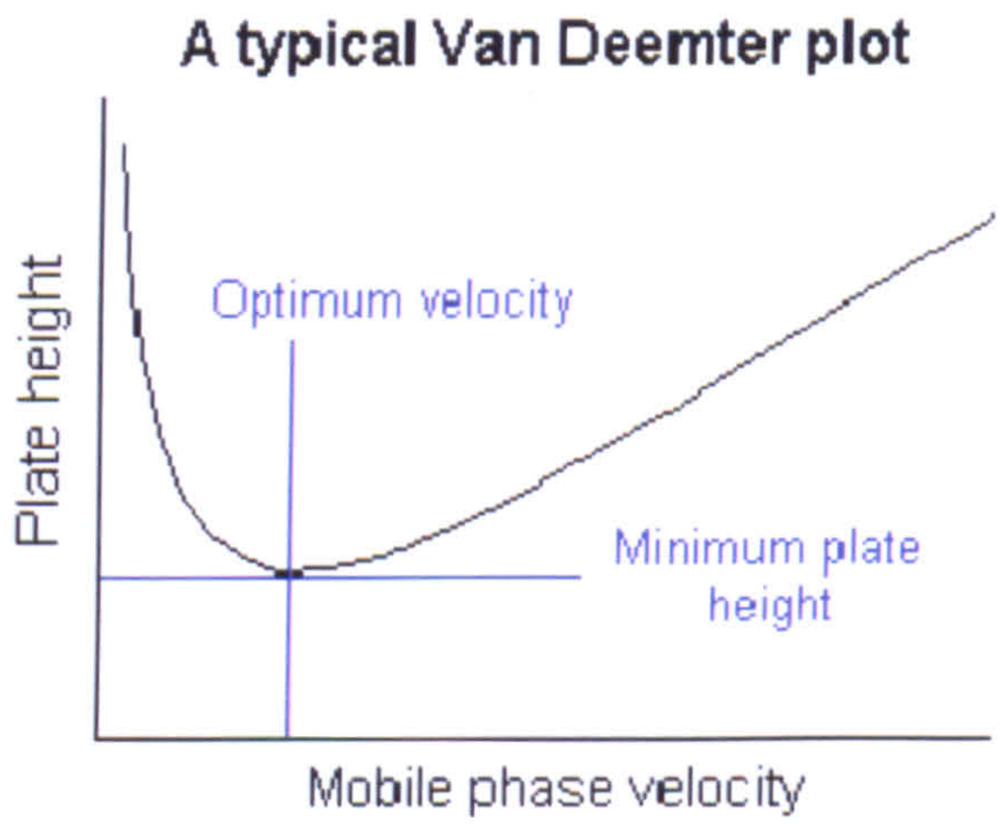


Figure 2.2 The Van Deemter equation predicts the minimal plate height and the optimal rate for a chromatographic separation (Figure adapted from [132]).

2.5.1 Affinity Chromatography

Affinity chromatography separates macromolecules through the use of a specific reversible interaction, such as that between a receptor and its ligand, antigen and antibody, or enzyme and substrate. The ligand is attached to a matrix and usually it is separated from it by a spacer. This technique is very useful as the first step of a protein purification, and allows a great capacity and good resolution. Starting from a huge volume, the protein will result in being concentrated and relatively pure. In

order to elute the protein, the interaction is weakened by a competitor or non-specific ligand, by changing the pH, the polarity or the ionic strength.

Proteins that have affinity for metallic ions can be separated by Immobilized Metal Ion Affinity Chromatography (IMAC). Some amino acids, like histidine or cysteine, can interact with ions at basic pH. The matrix is usually a polysaccharide, such as agarose, which prevents aspecific interactions. Chelating groups capable of binding ions are attached to the matrix and the ions used are Ni^{2+}, Zn^{2+}, Co^{2+}, Fe^{3+}, Cu^{2+} and Ca^{2+}, which carry partially empty d orbitals that can be occupied by electrons present in some amino acids (Figure 2.3). For example, the imidazole ring of histidine has a free electron pair at a neutral pH (pK_a histidine = 6.6). In this case, a high concentration of imidazole is used for the protein elution. Biological interaction can also be used to get rid of an unwanted protein. For example, it is possible to remove thrombin using a benzamidine column. The benzamidine, a competitive inhibitor of serine proteases, is covalently linked to a long spacer arm that is attached to the agarose matrix. This interaction can be broken at low pH.

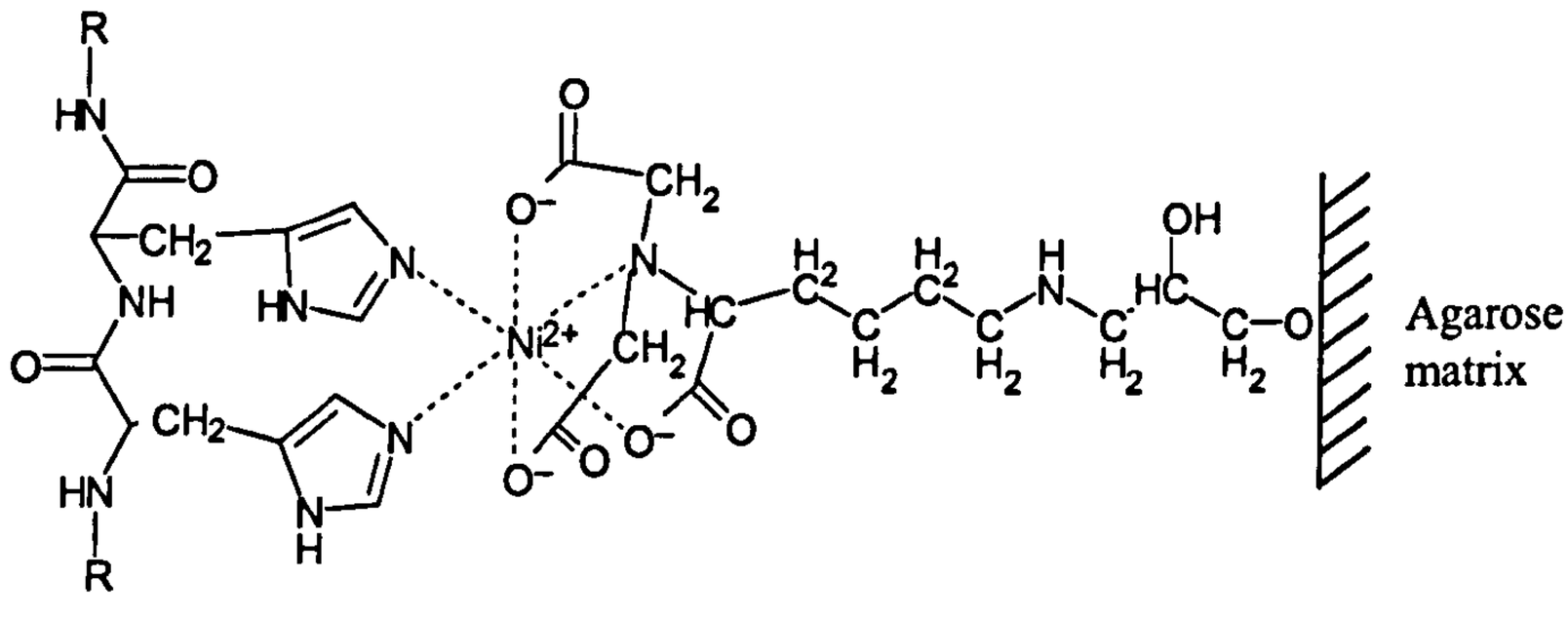


Figure 2.3 The structure of the IMAC resin used. The spacer carries nickel ions, which interact with two histidines (Figure adapted from the Amersham purification handbook).

2.5.2 Ion exchange chromatography

In ion exchange chromatography (IEC) macromolecules are separated according to their charge. The separation relies on a charge-charge interaction between the analyte and the charged matrix. IEC can be divided into: cation exchange chromatography, in which positive charges are exchanged with a negatively charged matrix, and anion exchange chromatography, where negative charges are binding to a positive matrix. In the first case sulphate derivatives (S-) or carboxylate derived ions resins (CM-) are used, while quaternary amine (Q-) or diethylaminoethane (DEAE-) cellulose resins are utilized in the second. Elution is achieved by increasing the ion strength. Ions from the mobile phase will interact with the matrix or with the protein, thus preventing an interaction with the stationary phase. Alternatively, elution can be achieved by altering the pH. This alters the charge of the amino acid lateral chains and therefore causes a change in the whole charge of the protein.

2.5.3 Size exclusion chromatography

Size exclusion chromatography (commonly known as gel filtration), allows the separation of different macromolecules according to their hydrodynamic volume. The difference from the other chromatographic methods relies on the fact that no chemical or biological interaction occurs. The stationary phase is a matrix composed of porous spherical particles that are chemically and physically inert. The particles can be made of silica or by organic polymers, and the pores have different sizes. These pores can either be channels through the bead, or depressions on the surface. The elution is isocratic, so the starting conditions do not change. Big molecules are eluted before smaller ones. The separation depends on their ability to enter the pores of the stationary phase. The molecules with a volume which is bigger than the pores will move faster than the ones that are smaller than the pores. It is possible to define an elution volume $V_e = V_0 + k_{av}V_i$, where V_0 or void volume, is the volume of the mobile phase surrounding the beads; V_i is the volume inside the beads, also called stationary phase volume V_s; and k_{av} is the distribution coefficient of the solute between the mobile and stationary phases:

$$k_{av} = \frac{V_e - V_0}{V_t - V_0} \text{ (where } V_t = V_o + V_i)$$

A completely retained molecule will have k_{av}= 1, and the volume of elution will coincide the entire value of the column: $V_e = V_0 + V_i$ (permeation limit). On the contrary, a molecule that is not retained will have $k_{av} = 0$ and $V_e = V_0$ (exclusion limit). It is possible to separate molecules that have elution volumes included between the two limits. In this range, there is a linear relation between the logarithm of the molecular weights and the elution volume. This relation allows to determine the molecular weight of an unknown sample starting from its elution volume and a calibration curve obtained with standard molecules. Finally, size-exclusion chromatography can also be used as a quick and simple method for buffer exchanging or desalting.

2.6 Electrophoresis

Electrophoresis, the migration of ions under the influence an electric field, is a common technique used to separate and identify single components from mixture of macromolecules, such as DNA, RNA, protein. This method allows to verify the presence of a target molecule, its degradation and purity level. Particles that are dispersed on a fluid carry an electrical surface charge. A molecule with a charge q under the electric field E is under the influence of the electric force $F_e = qE$, that drives the molecule towards the pole with opposite charge. The movement of the charged particles is also under the influence of a frictional force of opposite direction, $F_a = 6 \pi r \eta v$, that opposes the movement. r is the molecule hydrodynamic radius, η is the viscosity of the medium and v the velocity of the charge particles, that depends on both the viscosity and the shape and dimension of the molecule itself. When the two forces are the same, the molecule will move with a constant velocity $v = (E q) / (6 \pi r \eta)$. If E and η are constant, v depends on the ratio between the charge and the hydrodynamic radius ($v \propto q/r$). It is possible to define the electrophoretic mobility for a charged particle:

$$\mu = v / E = q / (6 \pi r \eta)$$

This property is used to separate different macromolecules that have a characteristic charge and mass ratio, and therefore migrate with a distinctive velocity under an electric field.

Gel electrophoresis is widely used with protein and nucleic acid. Porous gels are commonly made by polyacrylamide (polyacrylamide gel electrophoresis (PAGE)

gel) for protein or agarose for DNA. SDS-PAGE gels run in denaturing conditions. Sodium dodecyl sulphate (SDS) is an anionic detergent that denatures secondary and non disulfide linked tertiary structures. SDS binds with a constant ratio of 1.4 g of SDS for g of protein (about a molecule of SDS every 2 amino acids), conferring a negative charge to each protein in proportion to its mass. The charge of the lateral chain can therefore be ignored. All proteins will be transformed into elongated objects that present the same charge and mass ratio, but with a mobility that depends only on the molecule dimension. The biggest proteins will be delayed by the porous gel more than the smallest.

2.7 ITC

Isothermal Titration Calorimetry (ITC) is a biophysical technique that is used to study the thermodynamic parameters of interaction among small molecules. The ITC experiment gives a quantitative measure of the binding affinity (k_d), binding stoichiometry (n) and variation in enthalphy (ΔH) of the interaction between two molecules in solution. A VP-ITC unit was used for the experiment. It consists mainly of a spinning syringe for injecting and mixing the reagents and two chambers, a reference and a sample cell. Prior to addition of ligand, a constant power is applied to the reference cell. This directs a feedback circuit that activates a heater located on the sample cell. This feedback power is used to maintain the temperature equilibrium. The syringe containing the ligand is titrated into the cell containing the receptor. The temperature difference between the cells is recorded. For exothermic reactions, a negative change in the differential power (DP) occurs, since the heat evolved in the reaction can be used instead of the feedback power. The contrary happens for endothermic reactions. Since the DP has units of power (μCal/sec), it is possible to measure the ΔH by integrating each peak over time. The relation between released or adsorbed heat and the amount of binding that occurs is directly proportional. When the saturation occurs, the heat signal diminishes until the background heat of dilution can be observed.

For a ligand X binding to a single set of n identical sites on a macromolecule M ($MX_{n-1} + X = MX_n$), it is possible to define a single-site binding constant K:

$$K = \frac{[\text{filled sites}]}{[\text{empty sites}] \cdot [L]}$$

and the Gibbs free energy changes

$$\Delta G^0 = - R\ T\ \ln K = \Delta H^0 - T\ \Delta S^0$$

R is the gas constant and T is the absolute temperature, and ΔH^0 and ΔS^0 the changes in enthalpy and entropy for a single site binding in standard conditions. K, ΔH^0 and n are determined by non linear least square fit of calorimetric titration data, and ΔG^0 and ΔS^0 can be then quantified. The critical parameter that determines the shape of a binding isotherm is the unitless constant c:

$$c = K\ M_{tot}\ n$$

M_{tot} is the total concentration of macromolecule and n the stoichiometry parameter. Very tight bindings lead to very high c values, and the isotherm has a rectangular shape. Its height corresponds exactly to ΔH^0 and there is a sharp drop that identifies n. The shape of this curve remains constant when K varies but c remains above 5000. If c is reduced by decreasing M_{tot} the drop around n broadens and the intercept at the Y axis is lower than the real ΔH^0. At a very low initial M_{tot}, concentration the isotherm becomes similar to an horizontal line indicating a very weak binding (Figure 2.4).

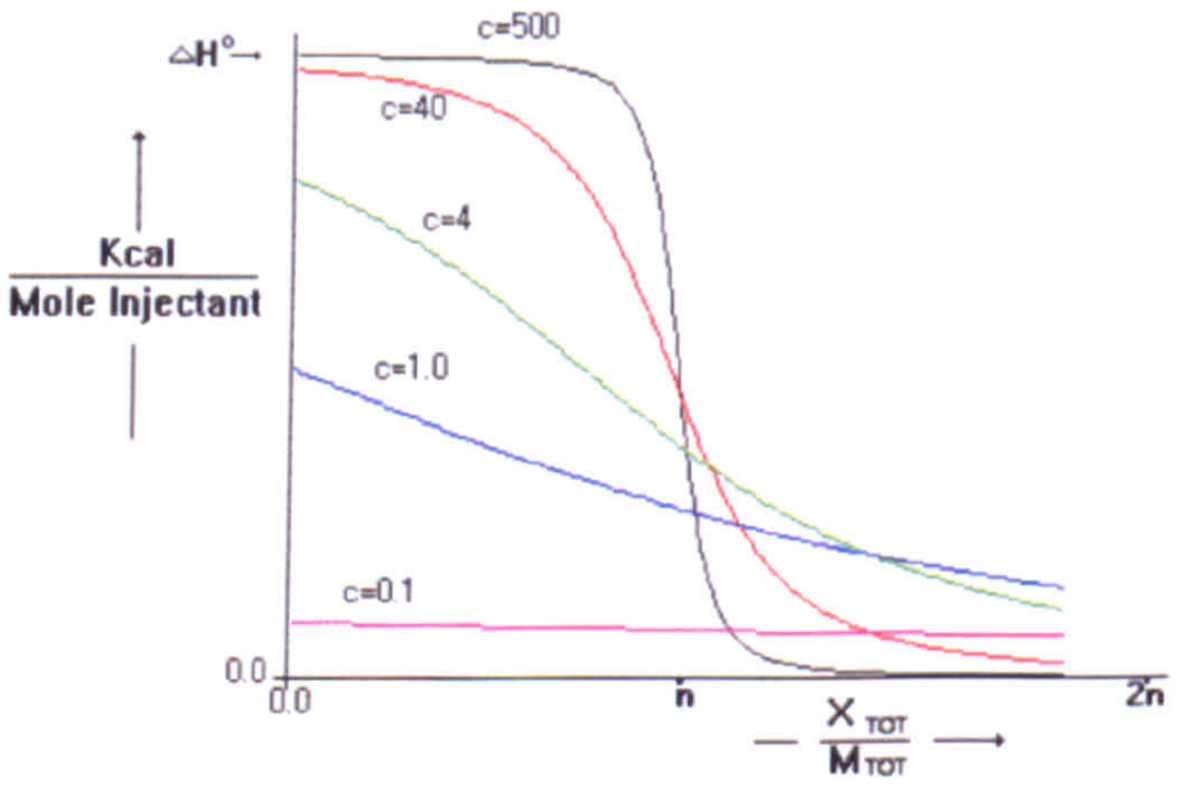


Figure 2.4 The ITC constant c. Values of C between 5 and 500 are ideal for measuring K with an ITC experiment (Figure adapted from the Microcal manual).

For the best results, ideally the concentration of ligand should be 8 times more than the protein concentration. The buffer in which the ligand is dissolved must be the same of that of protein. It is necessary to avoid heat effects caused by buffers mixing and control experiments must be carried out in order to determine the heat of dilution of ligand.

2.8 Protein crystallography

2.8.1 Protein crystallization

In order to solve a structure with the X-ray diffraction technique it is imperative to crystallize the protein as first step. The crystallization process is influenced by a lot of variables and is characterized by the nucleation process, the crystal growth, and the cessation of growth. The formation of a crystal depends on many physical and chemical parameters, specific of the reagents used, and on experimental variables that can determine its reproducibility, such as: the volumes used (0.5-2 μl); the process of mixing a protein solution and the precipitant; the temperature fluctuations; and the presence of additives. The protein must be very pure and characterized by a high level of homogeneity. The aqueous solvent is also essential for the conservation of the three-dimensional structure inside the crystal. Macromolecule crystals include a volume of solvent between 20% and 70% of the total volume of the crystal. The conventional methods of crystallization for small molecules, such as the total evaporation of the organic solvent or the drastic variation of temperature, will cause the crystal structure to collapse. The crystallization process must also be relatively slow. In this way, it is possible to obtain crystals with bigger dimensions and a higher level of regularity in the arrangement and orientation of the molecule inside them (low mosaicity).

For crystal growth, macromolecules must be in a supersaturation condition, a state of thermodynamic instability that returns to equilibrium as precipitate or as crystalline form. Supersaturation can be achieved by slow evaporation of the solvent or by changing the experimental variables. During crystal growth, it is possible to define three different steps, which happen at a different values of protein and precipitate concentration:

1. the nucleation phase: the spontaneous formation of the first nuclei is a statistical phenomenon whose probability increases the higher the level of supersaturation is. As the number of the nucleation centers increase, the level of supersaturation decreases. In addition, during the formation of the first aggregates, the formation of ordered nuclei can compete with the irreversible formation of amorphous aggregates, a process that can decrease the degree of supersaturation;

2. the growth phase, during which the crystal grows around the first protein nuclei. This phase needs a lower energy and a supersaturation lower than the nucleation phase;
3. the cessation of the growth. Several causes can determine the cessation of growth, such as the decrease in concentration of the crystalline solute to the point where the solid and the solution phases reach the equilibrium, growth defects, poisoning of the faces and ageing of the molecules.

The process of crystal formation depends on both the chemical properties of the protein and on its interaction with the solvent. The most common parameters that affect the protein solubility are:

- the temperature;
- the net charge: pH and pK_a of the buffer;
- the ionic strength: the salt concentration;
- the dielectric constant;
- the property of the precipitant.

These parameters can affect the solubility of the protein, but they can also affect the nucleation and growth phases. In solution, the stability of a macromolecule is controlled by balancing its interactions with the solvent and its intramolecular ones, which are responsible for its tertiary structure. Changing the pH of the solution will change the charge of some amino acid residues and therefore the net charge of the protein. At values near the pK_a, the solubility of charged residues can vary very quickly, affecting the overall charge and therefore the protein solubility. On the contrary, protein solubility is relatively stable to pH changes at pH values far from the pK_as. The ionic strength depends on the salt concentration and on the charge of each ion, and it affects the protein-buffer, protein-salt and salt-buffer interactions. Normally, when the salt concentration increases the protein solubility first increases (the salt ions interact with the protein charges), and then decreases. The protein solubility is also affected by the nature of the salt. The ability of salt to influence the solubility is summarized by the Hofmeister series. The relationship between solubility and temperature is an intrinsic property of a protein and cannot be predicted. Generally, the crystallization process is slower at lower temperatures. Finally, the polarity of the solvent is another parameter to consider. The polarity

can be reduced by adding a variable amount of an organic solvent, whilst avoiding high concentrations that can denature the protein in solution.

2.8.2 Vapour diffusion technique

Isothermal vapour diffusion is one of the most commonly used techniques for achieving the supersaturation condition that is necessary for the formation of crystals. A drop of protein solution, mixed with the precipitant solution, is equilibrated with a reservoir containing the same precipitant at a higher concentration. The two solutions are placed in a closed and isothermal system. The equilibration occurs via diffusion of the volatile species, like water or organic solvents, until the vapour pression of the drop and that of the reservoir are the same. If the process occurs with the evaporation of the water in the drop, the volume of the drop decreases while the protein concentrates till the supersaturation value. The solubility of the protein depends on many parameters, and a phase diagram can simplify the understanding (Figure 2.5).

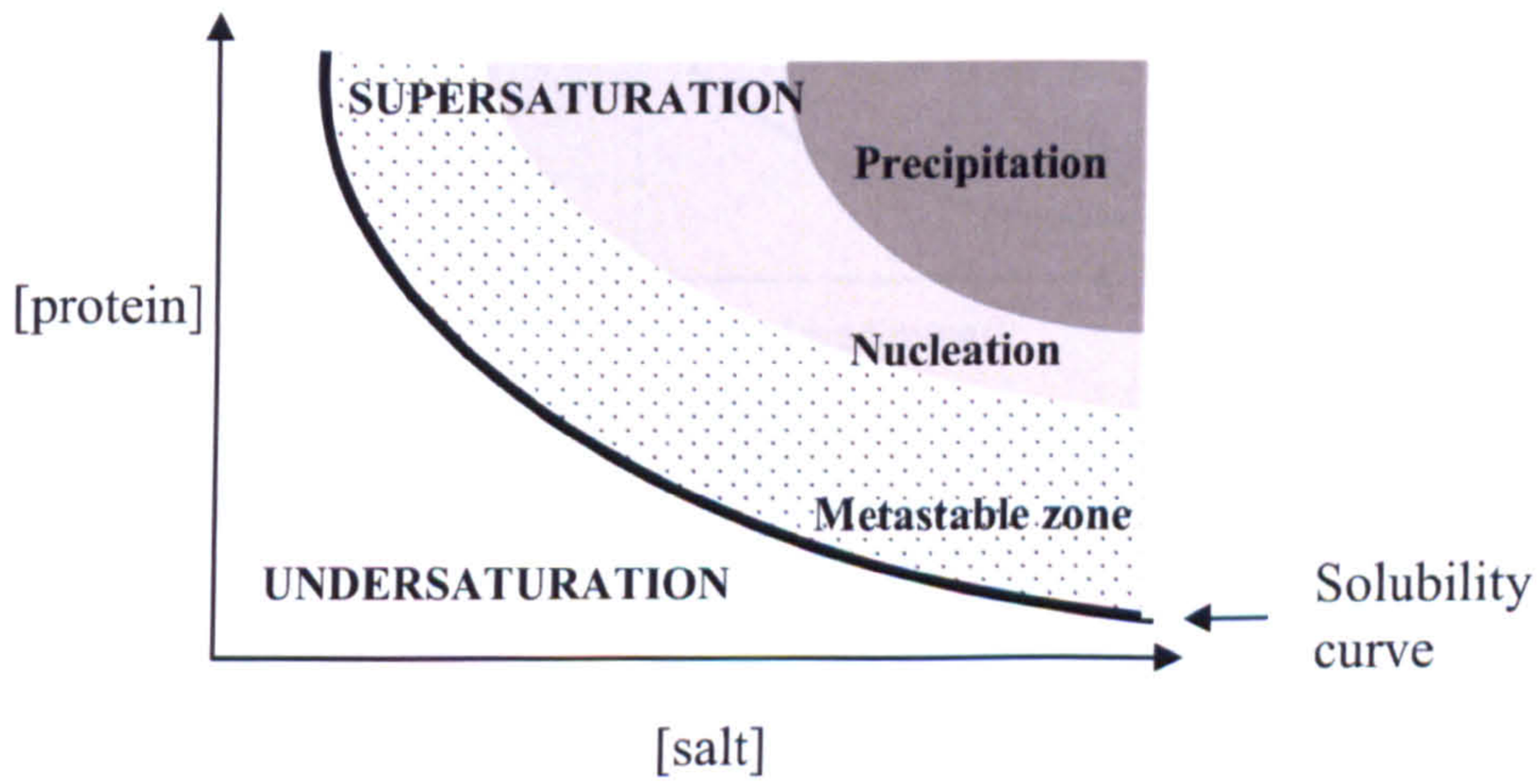


Figure 2.5 The solubility curve. The curve delimits the supersaturation and undersaturation zones.

The solubility curve defines the zones of undersaturation and supersaturation. Below the solubility curve the solution is not saturated, the system is thermodynamically stable and the protein cannot crystallize. On the contrary, in the supersaturation zone the concentration is higher than that present at the equilibrium. In this region is possible to distinguish the precipitation, nucleation and metastable

zones. The first one occurs at very high protein concentrations, and the insoluble molecules precipitate as an amorphous state. Nucleation requires a low supersaturation, while in the metastable region the critical supersaturation is not yet achieved, and spontaneous nucleation occurs only if induced by vibration or by the presence of additives. Usually, when the concentration of the precipitant in the reservoir is higher than that in the drop, the protein starts to concentrate from the not saturated A state (Figure 2.6), characterized by the concentration C_{ip}, to the supersaturated B state, at the C_{fp} concentration. If the nucleation zone is reached, the first crystals begin to appear (A → B in fig. 2.6 B), and the supersaturation is reduced. The protein that remains in solution will converge to the C state on the solubility curve.

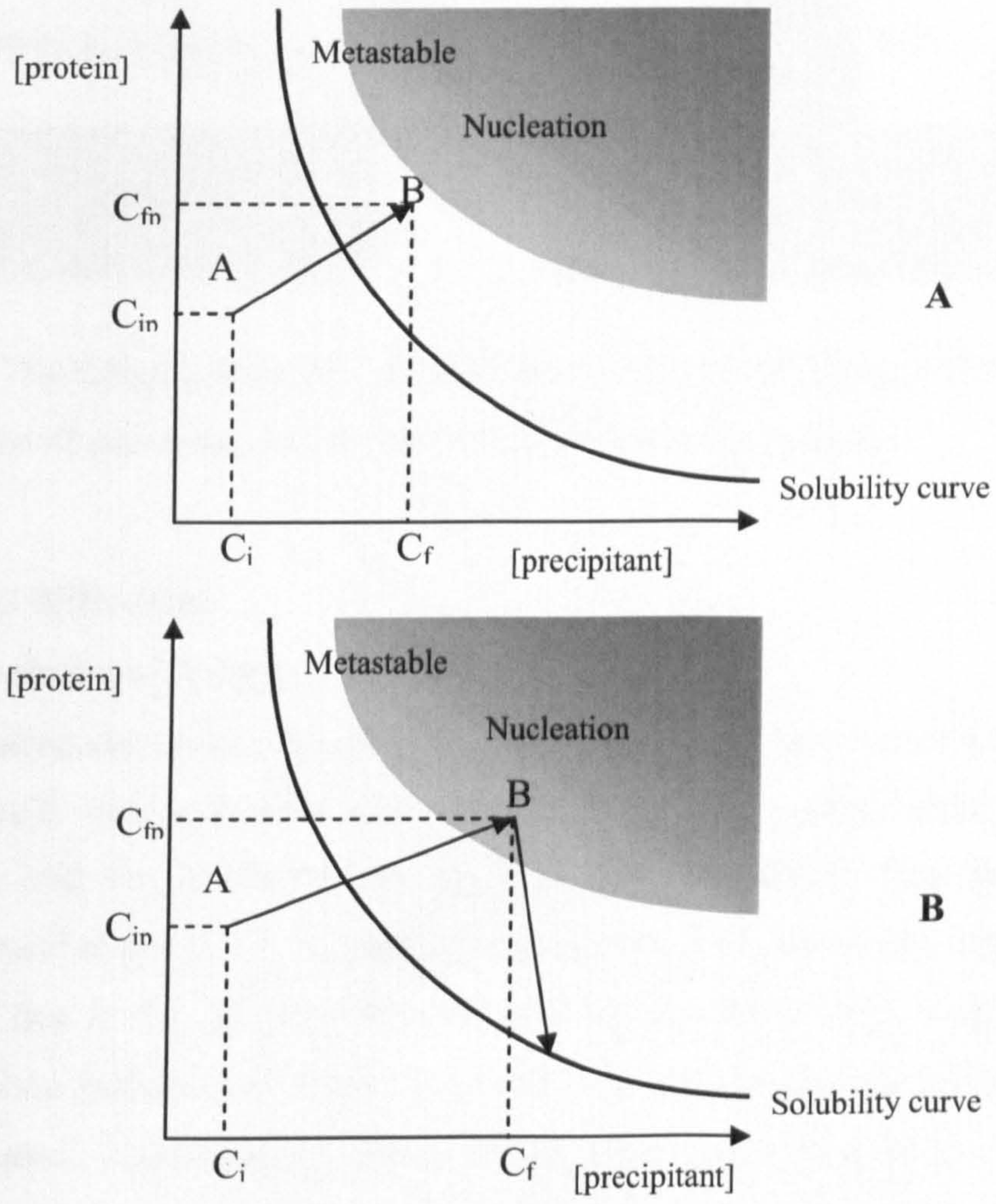


Figure 2.6 The nucleation zone. Solubility curve of a protein in function of the concentration of precipitant and of the protein concentration with the isothermic vapour diffusion. The crystal formation occurs when the nucleation zone is reached.

This condition can be achieved with either the hanging drop or the sitting drop methods (Figure 2.7). In the first case, the protein solution mixed with the precipitant agent is placed on a coverslip that is then turned over and placed above a well containing the reservoir at a higher concentration. With the second methods, the drop is placed on a bridge that is placed inside the well which is hermetically closed. In both cases the system is isolated with some grease or oil.

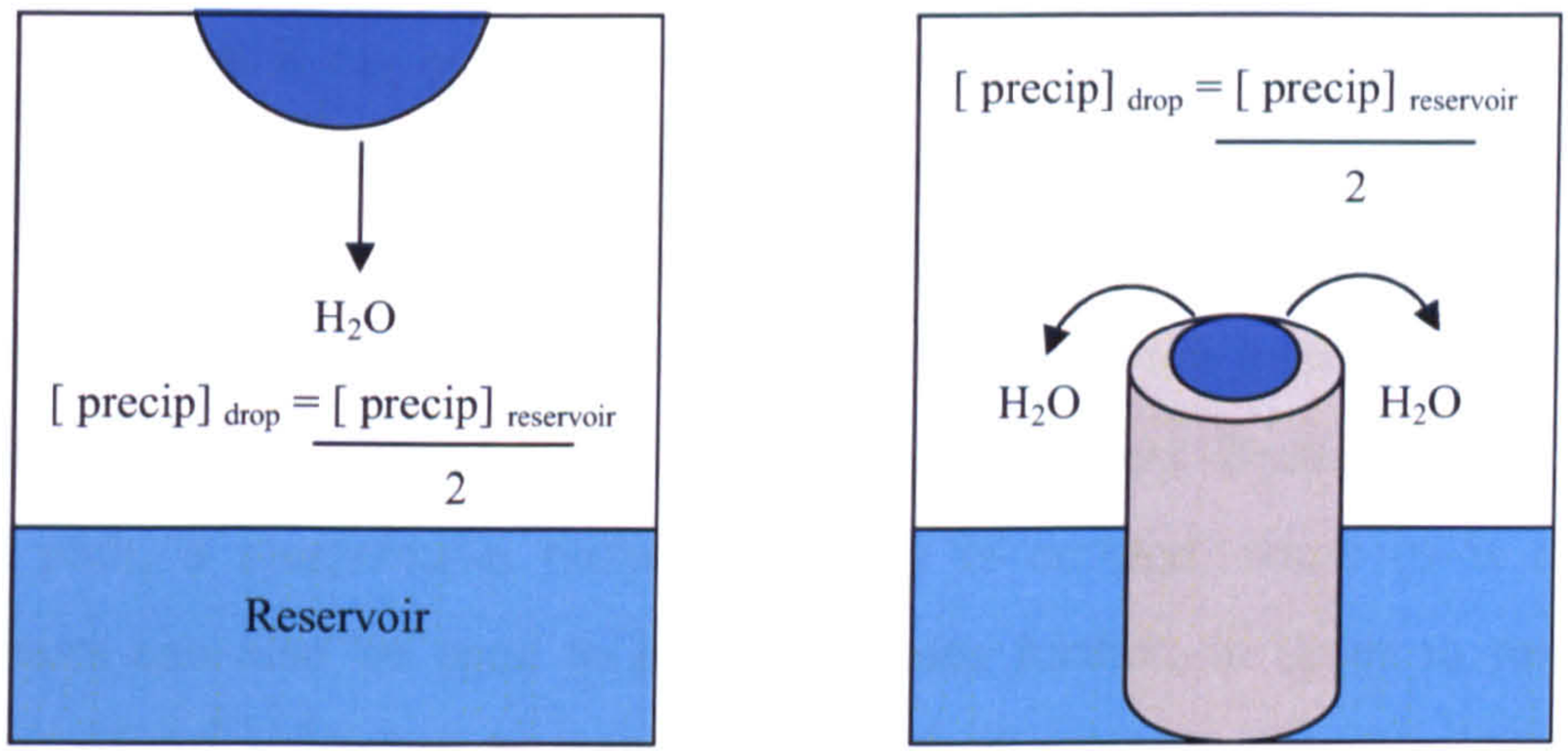


Figure 2.7 The hanging drop and sitting drop technique with vapour diffusion. The concentration of precipitant in the drop is lower than in the reservoir.

2.8.3 X-rays diffraction

2.8.3.1 The choice of X-rays

X-ray crystallography gives the possibility to determine the structure of a biological macromolecule, such as protein, DNA and RNA. An object must diffract the light to be seen, and the wavelength of the light must be smaller than the object. Individual bonded atoms 1.5 Å apart cannot be seen with the visible light, whose wavelength lies in the range of 400-700 nm. On the other hand, x-rays are an electromagnetic radiation of about 1 Å, and thus are suitable for diffraction of individual atoms. Unfortunately, lenses for an X-ray microscope do not exist. X-rays cannot be focused, and also most of the X-rays pass through the molecule without being diffracted. The problem can be solved by measuring the intensities and directions of the diffracted X-rays, and with the help of a computer that simulates a microscope lens. Protein crystals are ordered arrays containing a certain

number of identical molecules with identical orientation that the diffracted beams can add up, producing a strong unique detectable diffraction pattern.

There are two common sources of X-rays: rotating anodes and synchrotrons. In the first source a metal anode, typically copper, is bombarded with electrons of high potential (> 10 KeV) from a cathode, usually a tungsten filament. The anode rotates to allow cooling and a high current. These high energy electrons collide with the atoms in the anode, causing the displacement of electrons in lower orbitals. Electrons from higher orbitals drop into these empty orbitals, thus emitting their excess energy as X-ray photons. The emission spectrum of copper is characterized by two peaks, Cu-Kα (λ = 1.54 Å) and Cu-Kβ (λ = 1.39 Å), with wavelengths similar to those of individual bonded atoms. The other source of X-rays is the synchrotron, a particle accelerator ring where electrons are generated and forced to circulate at about the speed of light by powerful magnets. Because of the electrons curved path, a synchrotron radiation energy is emitted, some of it as X-rays. Undulators can also be used to bend the beam further, in order to magnify the intensity of the radiation. In comparison with the rotating anodes, synchrotrons allow a greater energy and the possibility of tuning wavelength selecting a wavelength from a comparatively large spectrum.

2.8.3.2 X-ray diffraction

Diffraction can be seen like a reflection (and for this reasons each spot is called reflection) from sets of equivalent parallel planes of atoms in a crystal, called Bragg or lattice planes. When X-rays interact with an ordered array like a protein crystal, the diffracted rays can interfere either constructively or destructively. Waves with the same phase will add up giving an amplitude that is the sum of the amplitudes of the individual scattered rays. On the contrary, out of phase rays interfere destructively resulting in an amplitude that is the difference between the individual rays.

Bragg's law shows that two waves diffract in phase from the planes (Figure 2.8) only if θ, the incident and reflection angle, meets the condition:

$$2\,d \sin\theta = n\,\lambda$$

where d is the interplanar distance, λ the wavelength, and n an integral number. R_2 travels the same distance as R_1, plus the added distance 2l. Using trigonometry, 2l

is equal to 2d sin θ (sine θ = BC/AB). If this distance is equal to an integral number of wavelengths, n λ, the two wavelengths will interfere constructively resulting in a diffracted beam.

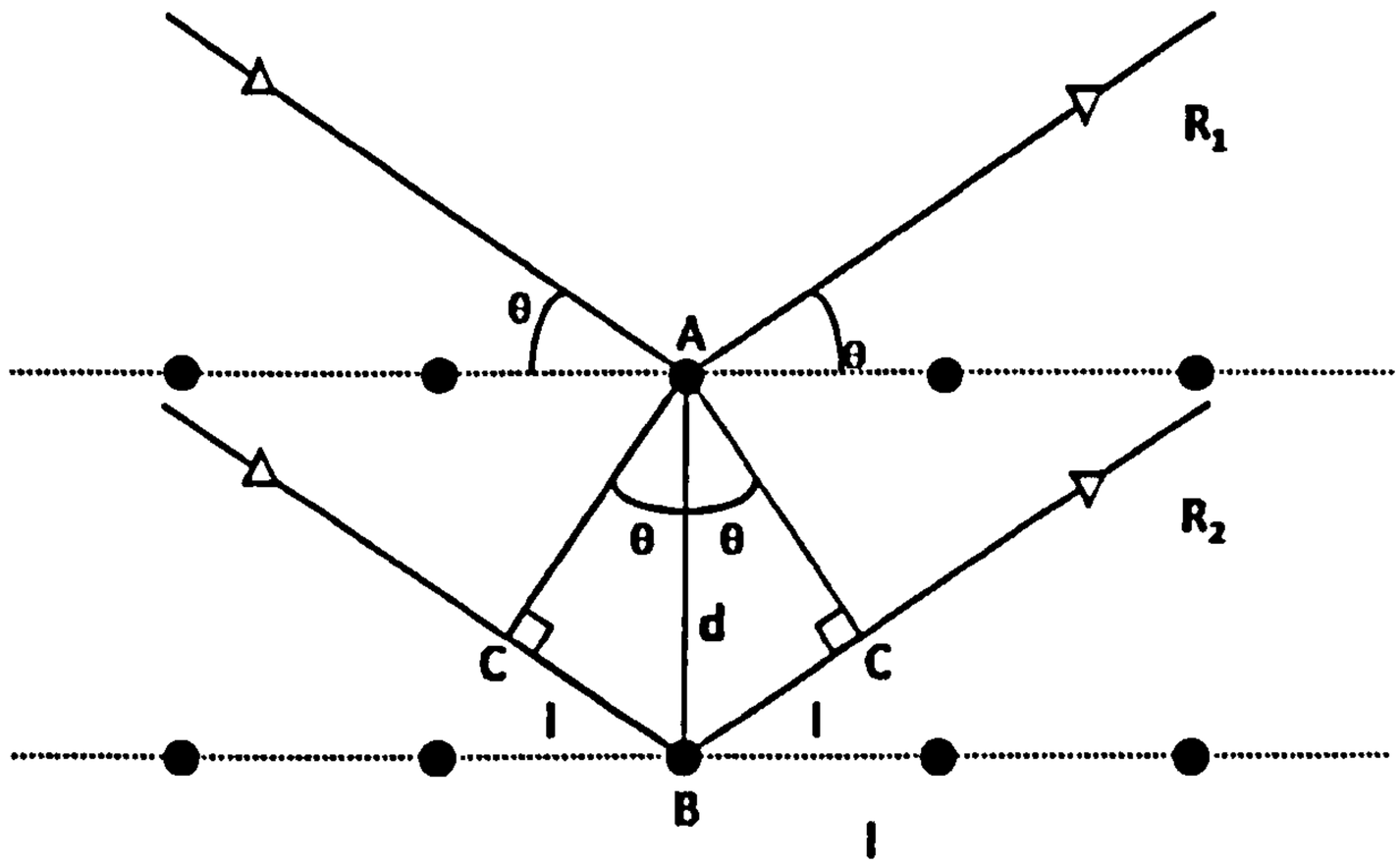


Figure 2.8 Condition that produces diffracted ray according to the Bragg's law.

In addition, if d has a value of 1, atoms on the planes will diffract in phase, while objects between the planes diffract out of phase. The angle of diffraction is also inversely related to the interplanar distance d, so for large unit cells (the smallest and simplest volume element that can reproduce the entire crystal by simple translation) the angle of diffraction will be small, producing more reflections that collide with a convenient angle from the incident beam than for a small unit cell.

The Bragg's law is also valid in reciprocal space. If a sphere of radius 1/λ and with a point located at the origin of the reciprocal lattice is drawn, all lattice points that make contact with the surface of this Ewald's sphere can satisfy Bragg's law and result in reflections (Figure 2.9). Rotating the crystal by a phi-axis during a diffraction experiment allows the collection of almost all the reflections. In other words, rotation of the reciprocal lattice allows sets of lattice points to make contact with the Ewald's sphere. The theoretical maximum number of reflections that is possible to measure is equal to the number of lattice points contained in a sphere of radius 2/ λ and centre at the origin of the reciprocal lattice, called the limiting sphere.

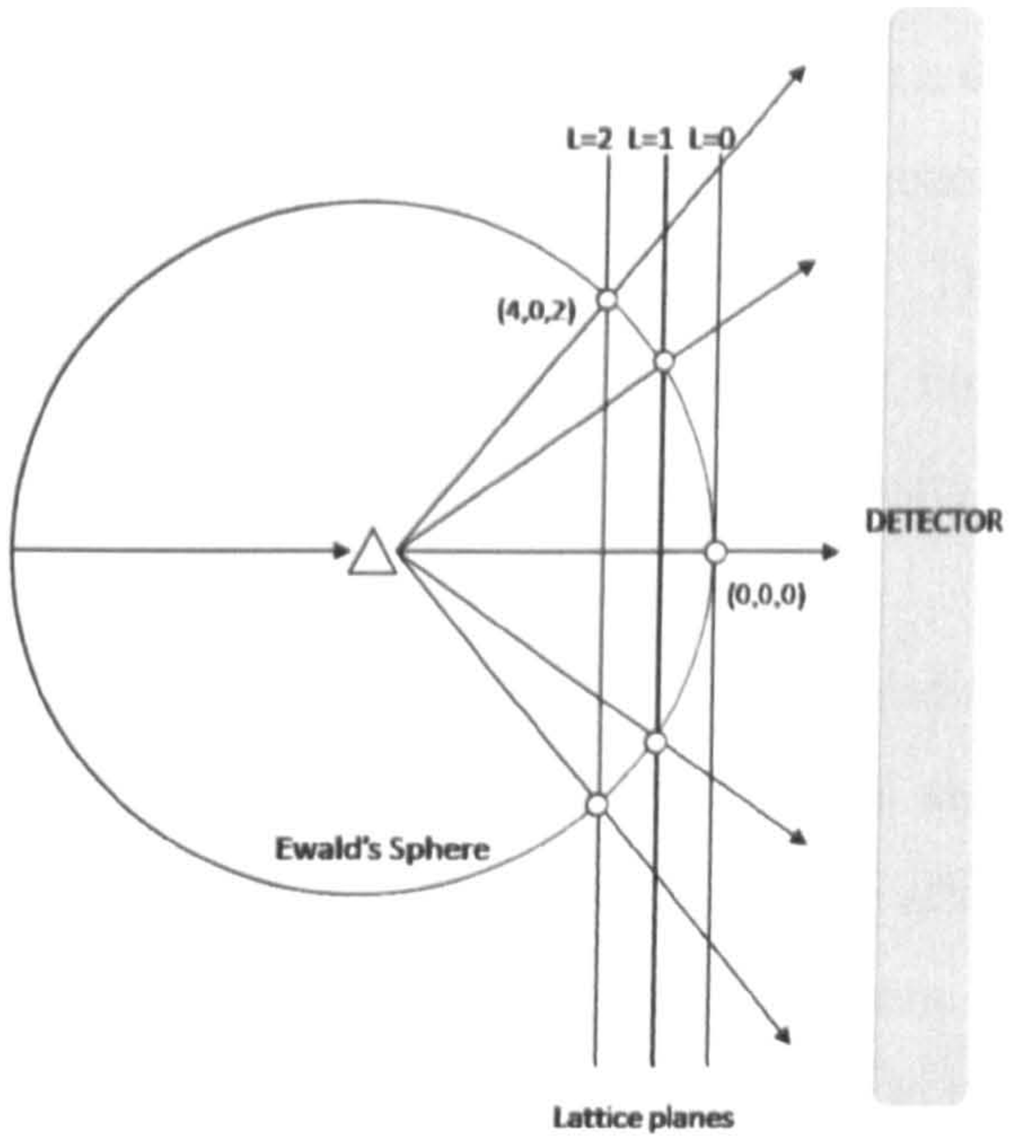


Figure 2.9 Ewald's sphere of reflection. Each lattice point making contact with the Ewald's sphere gives a diffraction spot. The sphere has radius $1/\lambda$ and one of its point is located at the origin of the reciprocal lattice (0,0,0).

2.8.3.3 Data collection, processing and reduction

The diffracted X-ray beams are measured by a detector that is placed in the path of the beam. Originally an X-ray sensitive film was used, while now an image plate device or a CCD detector is preferred. An image plate consists of a layer of phosphor crystals that cover the plate. The plate is scanned between exposures by a laser and the emissions from the excited phosphor layer are recorded. The emissions are proportional to the intensity of the x-rays that collide with the image plate at each diffraction spot. Solid-state electronic charge-coupled device (CCD) detectors are more recent, and eliminate the scans between each exposure, decreasing the time needed for the experiment. CCDs count the photons that are emitted from the phosphor layer.

The first diffraction images, usually taken 90° apart, are very useful to determine the best strategy necessary to obtain the best data from both a quantitatively and qualitatively point of view. These images can be autoindexed to determine the lattice type and the unit cell dimensions. Afterwards, it is possible to decide the optimal data collection strategy. Several programs utilize the cell dimensions and the spacegroup to determine the starting orientation of the crystal and the number of

frames to be acquired for the best completeness in the minimum time. The incoming diffraction images are then processed to generate a data file in which the intensity and position (hkl indices) of each reflection are recorded. Finally, all the data contained in the data file are scaled and merged. This compensates for fluctuations in the intensity of the beam and permits the combination of partial reflections, giving a final file containing the average intensity for each unique *hkl* position.

In addition, after scaling many of important statistical indicators regarding the quality of data become available. The ratio between signal and noise, or $I/\sigma I$, can indicate the highest resolution where the crystal diffraction is reliable and comparable to background noise or measurement errors. Usually, data are considered no longer reliable when this ratio drops below 2 in the highest resolution shell. The agreement between multiple measurements of the same reflections in a data-set is defined as R_{merge}:

$$R_{merge} = \frac{\sum_{hkl}\sum_{i}\left| I_i - \bar{I} \right|}{\sum_{hkl}\sum_{i} I_i}$$

where I_i is the ith intensity measurement of the reflection hkl, and $\bar{I}$ is the average intensity from multiple observations. R_{merge} values around 10% over the whole data-set and around 40% in the highest resolution shell are accepted. Redundancy describes the average number of observations acquired for each reflection, so the higher redundancy is, the better the data are. Finally, it is also important to consider the completeness of the dataset, which is the percentage of the calculated maximum number of reflections that has been collected. An incomplete dataset will result in zeros in the Fourier summation (chapter 2.8.3.4) that can introduce noise.

2.8.3.4 The Fourier transform

As a consequence of the X-ray diffraction experiment, a series of diffraction images are collected from the crystal. The aim of a crystallographic experiment is to calculate a map of the distribution of electrons in the molecule, the electron density map. The relationship between this electron density and the diffraction pattern can be described with the Fourier transform. If the electron density is seen like a mathematical function, the diffraction pattern is the Fourier transform of this

function. The structure factor F_{hkl} describes a diffracted ray that produces a reflection to the detector. The atomic structure factor f_{hkl} describes the contribution of a single atom *j* for the reflection *hkl*:

$$f_{hkl} = f_j e^{2\pi i(hx_j+ky_j+lz_j)}$$

f_j is the scattering factor of atom j, a mathematical function that describes the scattering of X-rays by an atom. This scattering function can be determined assuming that the electron density of each atom is a discrete and a simple sphere. For a unit cell containing *n* atoms, the structure factor F_{hkl}, describing all the diffractive contribution of all atoms for the reflection *hkl*, can be defined as a Fourier series in which each term is the contribution of one atom, treated as a simple sphere of electron density:

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hx_j+ky_j+lz_j)}$$

Alternatively, the structure factor F_{hkl} can be described in terms of sum of electron density contributions at different positions *xyz* within the unit cell:

$$F_{hkl} = \int_V \rho(x,y,z) e^{2\pi i(hx+ky+lz)} dV$$

where V is the total volume of the unit cell. The Fourier transform is a reversible function. So, the Fourier transform of the diffraction pattern results in the description of the electron density of all atoms within the unit cell. This Fourier transform relates the real-space electron density map to the reciprocal space structure factors:

$$\rho(x,y,z) = \frac{1}{V}\sum_h\sum_k\sum_l F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

In order to build the electron density $\rho(x,y,z)$, it is necessary to construct a fourier series from the structure factors.

2.8.3.5 The phase problem

The previous equation can be rearranged in this way:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F_{hkl}\right| e^{-2\pi i(hx+ky+lz-\alpha'_{hkl})}$$

F_{hkl} is in fact a periodic function, and can be described with an amplitude $\left|F_{hkl}\right|$, a frequency, and a phase α. The frequency is that of the X-ray source, the amplitude is proportional to the square root of the intensity I_{hkl}, but the phase is not directly measurable from the reflection intensity measure. That is the origin of the phase problem. Currently three main methods are available to overcome the phase problem. The heavy atom method (or multiple isomorphous replacement, MIR) relies on the introduction of heavy atoms, such as silver, gold etc, into the protein crystal. These heavy atoms, through their greater scattering factors, will dominate the diffraction. Using a Patterson function it is possible to determine their relative position. The Patterson function is a Fourier series calculated using only intensities and no phases:

$$P(u,v,w) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F_{hkl}\right|^2 e^{-2\pi i(hu+kv+lw)}$$

where u,v,w are coordinates in the Patterson map. Once the positions of the heavy atoms are determined, their contribution to the phase angles of each reflection can be calculated. The structure factor for an isomorphous derivate (F_{PH}) is equal to the sum of the structure factors of the native protein (F_P) and the heavy atoms (F_H). The multiple wavelength anomalous diffraction (MAD) relies on the anomalous scattering of some heavy atoms, like iron or selenium, at a specific wavelength. It is possible to introduce these atoms in the same way as in the MIR, or to heterogously express the protein in the presence of selenium methionine, using *E. coli* strains auxotrophic for this amino acid. It is possible to locate these atoms following the same MIR principle: anomalously diffracting atoms behave in a different way from the other atoms in the protein, absorbing X-rays at specific wavelengths. The phase can be calculated following the same MIR idea. In addition, MAD has the advantage that anomalous atoms scatter differently at different wavelengths, so only a single crystal is needed to collect the native and the derivate set of data.

Finally, the molecular replacement method (MR) relies on the availability of a structure with sufficient structural similarity to the unknown protein. Other direct techniques for determining the phase are also available. They rely on computational manipulation of the diffracted data without any additional information. These direct methods are extensively used in the crystallography of small molecules, and can be used also for macromolecules, but they must diffract at atomic resolution and this approach is in continuous development.

2.8.3.6 Molecular replacement

This method for phase determination relies on initial phases generated from a model with sufficient sequence and, more importantly, structural similarity:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}\left|F_{hkl}^{new}\right| e^{-2\pi i\left(hx+ky+lz-\alpha_{hkl}^{'model}\right)}$$

where the structure factor amplitudes are those of the new protein, and the phases those of the model. The search model is chosen based on a number of factors, including sequence and suspected structural similarity, that is essential for molecular replacement. When the unknown protein and the search model share a sequence identity of more than 30%, the probability they share also a similar structure is very high. In addition, if there is a sequence dissimilarity, then the difference in the structure is more restricted. This method calculates the correlation of the search model Patterson vectors to those of the unknown molecule, assuming that two molecules with a similar structure display similar Patterson functions. The model must be placed in the correct position within the unit cell. To perform this, a six dimensional search, comprising the translational xyz values and the rotational $\alpha\beta\gamma$ values, is carried out. The rotation corresponds to the rotational function, and is performed before the translational search because the Patterson map is independent of the position of an object within the unit cell as long as its orientation is maintained. The rotation is obtained by rotating one of the two Patterson maps, aiming to the maximum overlap. In addition, the Patterson map is determined mainly by strong reflections, so weak high resolution data are not used in this process. The next step is to locate the correct position, that is, translations in x,y,z

in the unit cell for the molecule. This step is carried out in real space. For each attempt, the structure factors are calculated and compared to the experimental ones. The R-factor:

$$R = \frac{\sum_{hkl} \left| |F(obs)| - |F(calc)| \right|}{\sum_{hkl} |F(obs)|}$$

and the correlation coefficient C:

$$C = \frac{\sum_{hkl} \left(|F(obs)|^2 - \overline{|F(obs)|^2} \right) \times \left(|F(calc)|^2 - \overline{|F(calc)|^2} \right)}{\left[\sum_{hkl} \left(|F(obs)|^2 - \overline{|F(obs)|^2} \right)^2 \times \sum_{hkl} \left(|F(calc)|^2 - \overline{|F(calc)|^2} \right)^2 \right]^{1/2}}$$

describe the difference between the calculated structure factors amplitude F(calc) and those observed F(obs). The best position solutions are chosen on the basis of the correlation coefficient and R-factor values, and also on the difference between the best correlation coefficient and the following one. Correct solutions can be identified because they will not cause clashes between a molecule and its symmetry related ones. The molecular replacement usually finishes with a rigid body refinement in order to improve the amplitude structure factor fits and the phase estimate.

2.8.3.7 Refinement and model validation

Structure factors calculated from the model are employed to calculate the electron density map. The electron density is usually displayed as both Fo-Fc difference map (structure factors observed minus those calculated) and 2Fo-Fc maps, with the former highlighting the difference between the calculated and observed data, and the latter giving more weight to the observed amplitudes. The aim of the refinement is for the model to describe the experimental data completely, that is, Fo to be equal to Fc. The Fo-Fc difference map is very useful to assess the progress of refinement, producing positive or negative peaks where Fo differs from Fc. The 2Fc-Fo map gives information about the shape of the structure and it is often in the form of a continuous molecular surface. This map is very useful for the construction of the

new model. After the first map is created, the model building can start. This is a process where the model is fit into the observed data. In addition to manual model building, a computational refinement is carried out, aiming to reduce as much as possible the R factor, implying an increase in the agreement between the model and the real structure. In order to achieve this, atom positions are changed to obtain the best density map fit and the temperature B factor. Software like *Refmac* [133] applies the principle of maximum likelihood, which is based on the fact that the probability of observing a particular set of observation should be very high. When changes are applied to the model, in the attempt to bring it closer to the actual model, there is an increment in the probability of observing the real structure. The principal problem with refinement is the fact that the number of observations is not very much superior to the number of parameters, causing a small over-determination of the structure. Because of this, it is necessary to either reduce the number of parameters or effectively increase the number of observations. Therefore, to increase the structure determination, new observations in the form of geometrical restraints, which keep bond lengths, bond angles and atomic contacts in the ranges typically seen in very high resolution structures, are added. Finally, the refinement of a structure is followed by different markers, like the R-factor, that is however subjected to bias by the refinement process. So, the unbiased R_{free} was introduced. It is calculated in the same way as the R-factor, but it refers to a set of reflections that are not part of the refinement process. These reflections are normally between 2-10% of the total reflections, picked randomly to minimize any sort of errors. Besides to the monitoring of these errors, the density map inspection is always a very useful tool to obtain a better model. The atomic model must agree with the electron density map. The analysis of the Ramachandran plot is also useful for checking the geometry and stereochemistry of the model.

2.8.3.8 Density modification

The first calculated map may have an insufficient quality or may be highly biased. Cycles of density modifications, like solvent flattening, non crystallographic symmetry (NCS) averaging and histogram matches can be used to obtain better phases and therefore electron maps. Solvent flattening smooths the map eliminating small spikes and random fluctuations in density, and defines the boundaries

between solvent and protein. For example, the unit cell can be divided into a grid of regularly spaced points. At each points, the value of ρ (x,y,z) in the Fo map is evaluated. If ρ is negative, it is reassigned a value of zero, while if it is positive it is assigned a value equal to the average value of ρ within a certain distance. Therefore, the map is divided into a region of relatively high and low density: the protein and the solvent regions.

The strength of the signal can also be increased by averaging the signals from identical molecules but having a different orientation in the unit cell. These molecules may be related by symmetry elements that are not aligned with symmetry elements that apply across the entire unit cell (crystallographic symmetry elements). By knowing the location and type of non crystallographic symmetry elements, it is possible to add the signal of these set of elements together, improving the signal to noise ratio. This modification technique can be successful with two-, three- and five-fold rotation axes.

A density histogram is a probability distribution of values of the electron density sampled at regular intervals throughout the three-dimensional map, and density histograms of protein structures have a characteristic form. In histogram matching, the electron density distribution of the map is compared to that of a standard map at similar resolution, and then it is modified trying to match the standard map in order to reduce possible errors. When all possible improvements have been done to the electron density map, the map can be back-transformed to give more accurate phases than those obtained imply from the atoms of the model alone.

Chapter 3

Materials and Protocols

3.1 Media

3.1.1 Luria-Bertani (LB) liquid medium

A pouch provided by MP Biomedical (10 g tryptone, 5g yeast extract, 10 g NaCl) was dissolved in 1 L of ultrapure water and then autoclaved at 121°C for 15 minutes. The appropriate antibiotic (carbenicillin (50 μg/mL) or kanamycin (50 μg/mL), Apollo Scientific) was then added prior to inoculation.

3.1.2 LB plates

100 mL LB with the addition of agar (1% w/v), provided by the Department of Cellular and Molecular Medicine, were boiled until the melting point was reached, then let to cool down to approximately 40°C. The appropriate antibiotic (carbenicillin (50 μg/mL) or kanamycin (50 μg/mL)) was then added and about 20-25 mL of medium were poured in a Petri dish (Sterilin), in a sterile environment. The agar was left to solidify again at room temperature and then the plate was stored at 4°C.

3.2 Vectors

- pET-11c-ACP (kindly donated by Dr Matt Crump and Dr John Crosby, Department of Chemistry) carries the ACP gene and the ampecillin resistance;
- pET-15b-AcpS (kindly donated by Dr Matt Crump and Dr John Crosby, Department of Chemistry), carries the AcpS gene with the addition of a N-terminal hexa-histidine tag coding region and the ampicillin resistance gene. AcpS H110A and D111A mutants were also cloned into this vector;

3.3 Host strain

- *E. coli* BL21 Star (DE3) One Shot (Invitrogen) competent cells. Genotype: F- ompT hsdSB (rB-mB) gal dcm rne131 (DE3)).

3.4 Molecular biology

3.4.1 Agarose gel electrophoresis

1g of agarose (Invitrogen) was dissolved in 100 mL TAE buffer (40 mM Tris acetate, 1 mM EDTA) and heated at 800W for 1-2 minutes with a microwave, until the solution was clear. The solution was left to cool down to about 40°C before adding ethidium bromide (Invitrogen) to a final concentration of 0.5 μg/mL. The solution was then poured into the plate and left to set with the comb in place. 5X sample running buffer (Bioline) was added to the DNA samples to give a final concentration of 1X. 5μL Hyperladder I markers (Bioline) were used as a molecular marker. The gel was run in TAE buffer at 100 V for 90 minutes, and then visualized under UV light.

3.4.2 DNA miniprep

10 mL *E. coli* liquid culture, with the appropriate antibiotic, was grown overnight at 37°C. The bacterial pellet was collected by centrifugation and plasmids were extracted using the QIAprep Spin Miniprep Kit (Qiagen) according to the protocol provided by the manufacturer. The DNA was stored at -80°C until required.

3.4.3 Restriction enzyme digest

- ACP restriction sites: *NdeI* and *BamHI*;
- AcpS, H110A and D111A mutants restriction sites: *NdeI* and *BamHI*;

The digest solution (2.5 μL of DNA, 9.2 μL of deionized water, 0.8 μL of BSA, 1.5 μL of buffer and 1μL for each appropriate restriction enzyme) was left at 37°C for 75 minutes. The result was checked by running an agarose gel. Restriction enzymes were bought from Roche.

3.4.4 Site-directed mutagenesis for AcpS mutants

AcpS-pET15b was first sequenced (10 μL at a concentration of 100 ng/μL, service provided by Cogenics Lark Inc.). Then, primers were designed to introduce the single mutation H110A or D111A and synthetized by MWG. The primers are described in Table 3.1. The site-directed mutagenesis was performed using the QuickChange II Site-Directed Mutagenesis Kit (Stratagene), following the suggested protocol. Various concentrations of 25, 50, 100 and 200 ng of DNA

template were tested, with 125 ng of each primer. 100 ng of DNA template gave the best results. PfuUltra HF DNA polymerase was used. In the PCR mixture, 1 μL of DMSO (for a final concentration of 20 μM) was added to avoid the formation of CG hairpins; this step turned out to be essential for the success of the experiment. As suggested, 16 cycles were chosen for the single amino acid substitution. After *DpnI* digestion, the DNA was transformed into XL1-Blue Supercompetent Cells. The higher number of colonies was obtained using 3 μL (for D111A) and 2μL (for H110A) of DNA. After extraction using the miniprep, the products of the mutagenesis were firstly subjected to restriction enzyme digest, and then the restriction products were checked in an agarose gel to see if they were of the appropriate size. Finally, they were sent to Cogenics Lark Inc. for sequencing.

Primer	Sequence (5' to 3')
D111A forward	TGG CAC GTG TCG CTG AGC CAC **GCC** GCG GGG ATC GCG TCG GC
D111A reverse	GC CGA CGC GAT CCC CGC **GGC** GTG GCT CAG CGA CAC GTG CCA
H110A forward	CAC GTG TCG CTG AGC **GCC** GAC GCG GGG ATC GCG TC
H110A reverse	GA CGC GAT CCC CGC GTC **GGC** GCT CAG CGA CAC GTG

Table 3.1 Primers used for PCR in AcpS site-directed mutagenesis. The bases in bold and underlined correspond to the mutation sites from aspartate (CAC) or histidine (GAC) coding region to alanine (GGC) coding region.

3.4.5 Heat-shock transformation of plasmid DNA into competent *E. coli*

1μL of DNA was transferred into 25 mL of supercompetent cells (Novagen). The solution was left in ice for 5 minutes, transferred into a water bath at 42°C for exactly 30 seconds, and put back in ice for 2 minutes. 125 μL of SOC medium were added and the culture was placed in the incubator at 3°C, 220 rpm, for 1 hour. 20 and 40 μL were then spread into Petri dishes, containing agar medium and the appropriate antibiotic, and were left overnight at 37°C. To test whether the transformation had been successful, 4 single colonies from each Petri dish were chosen and spread again in a new plate. 2 colonies from each plate were finally chosen and their DNA extracted.

3.4.6 Glycerol stock

1 colony was transferred into 10 ml LB medium with the appropriate antibiotic and left in an incubator at 37°C, 225 rpm, overnight. Then, 700 µL from each of the 10 mL culture were gently mixed with 50% glycerol in PBS (300 µL). Each tube was then frozen in liquid nitrogen and stored at -80°C.

3.5 Protein expression

A single colony from an agar plate (obtained from the glycerol stock) was transferred, using a sterile toothpick, into 100 mL of LB medium, with the appropriate antibiotic, in a 250 ml flask to set up an overnight culture (ONC) at 37°C under stirring (220 rpm). 500 mL or 1 L LB broth with the appropriate antibiotic was placed in a 2.5 L conical flask. 40 mL of the ONC were added, and the culture incubated at 37°C, 220 rpm. When the optical density of the culture reached about 0.6 (measured using spectroscopy at 600 nm with 1mL culture and using LB broth as a blank), IPTG was added to result in a final concentration of 1 mM. The culture was then incubated at 30°C (32°C for ACP) for 4-5 hours.

3.6 Protein purification

3.6.1 Buffers

All buffers were filtered (0.22 µm Minisart filters, Sartorius) prior to use. Buffers for gel filtration chromatography were also degassed under reduced pressure. All pH meter measurements were undertaken using a S20 Seven Easy pH meter (Mettler Toledo). Chemicals were bought from Sigma-Aldrich, BDH or Fisher Scientific, unless otherwise stated.

3.6.2 Cell harvest and bacterial lysis

Cells were harvested by centrifugation at 5000g, 4°C for 20 minutes. The supernatant was decanted off and the cell pellet was resuspended in a small volume of PBS, then centrifuged at 18000g, 4°C for other 20 minutes. The resulting pellet was then frozen at -80°C until required. The pellet was resuspended in about 30 mL of appropriate resuspending buffer (described in Table 3.2), with the addition of 0.1 mg/mL lysozyme and a tablet of protease inhibitors (Complete EDTA-free, Roche). The solution was then sonicated (VibraCell sonicator, Sonics) on ice, using the

macrotip at 70% amplitude, for 25 seconds. After a pause of 25 seconds, the sonication was repeated for seven/eight times, or until the solution was less viscous. The resulting solution was then centrifuged at 18000g for 20 minutes at 4°C. The pellet was discarded and the supernatant containing the soluble protein retained for the purification steps.

Protein	Resuspending buffer
ACP	50 mM Tris-HCl pH 8, 2 mM DTT
AcpS	50 mM Tris-HCl pH 8

Table 3.2 Lysis buffers used for each different protein.

3.6.3 Affinity chromatography

The affinity chromatography steps were performed using an FPLC machine (Akta Prime, GE Healthcare) and a HisTrap HP 5 mL column (Amersham). The flow rate was 1 mL/min. The Ni^{2+} affinity column was firstly equilibrated with the appropriate equilibrating buffer. The clarified supernatant (chapter 3.6.2) was filtered with 0.22 μm filters and then loaded into the column, that wash washed with a minimum of 3 column volumes of equilibrating buffer. The protein was eluted with a gradient of 70% elution buffer B in 50 mL. To check if other proteins were still bound to the resin, the gradient was then taken to 100% elution buffer B in 3 column volumes and maintained for another column volume. The resulting fractions were pooled and concentrated for further purification steps. Equilibration and elution buffers are described in Table 3.3.

Protein	Equilibrating buffer	Elution buffer B
AcpS	50 mM Tris-HCl pH 8, 5 mM imidazole, 10 mM $MgCl_2$, 500 mM NaCl	50 mM Tris-HCl pH 8, 1 M imidazole, 10 mM $MgCl_2$, 500 mM NaCl

Table 3.3 Equilibration and elution buffers used with the Ni^{2+} affinity column.

3.6.4 Ion exchange chromatography

The ion exchange purification was performed using a Q-Sepharose HiLoad 16/60 column (Amersham) and a flow rate of 2 mL/min. The Q-Sepharose anion-exchange column was firstly equilibrated with the appropriate buffer, then the protein solution was loaded and eluted with the gradient of salt (0-100% ion exchange buffer B over 15 column volumes). 2 mL fractions were collected. The fractions containing protein were identified using SDS-PAGE gel and concentrated for the following purification steps. The buffers used are summarized in Table 3.4.

Protein	Ion exchange buffer A	Ion exchange buffer B
ACP	50 mM Tris-HCl pH 8, 5 mM β-mercaptoethanol, 10 mM $MgCl_2$	50 mM Tris-HCl pH 8, 5 mM β-mercaptoethanol, 10 mM $MgCl_2$, 1 M NaCl

Table 3.4 Ion exchange buffers used with the ion exchange chromatography.

3.6.5 Size exclusion chromatography (gel filtration)

The previous fractions containing protein were pooled and concentrated to 500 µL or 1 mL. The column was initially washed with the appropriate equilibrating buffer with a flow rate of 0.5 mL/min. The protein solution was then loaded and fractions collected every 1 mL. Fractions were monitored using the online UV monitor measuring absorbance at 280 nm, by running samples on SDS-PAGE gels. The different buffers used are described in Table 3.5.

Protein	Equilibrating buffer	Column used (Amersham)
ACP	50 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 300 mM NaCl, 5 mM DTT	HiLoad 16/60 Superdex 75
AcpS	50 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 300 mM NaCl	HiLoad 16/60 Superdex 75 / Superdex 75 10/300 GL

Table 3.5 Gel filtration buffers used for gel filtration purification.

3.6.6 Thrombin cleavage of wild type his-AcpS and purification of untagged AcpS

The concentration of protein was measured by UV absorbance and Bradford assay. 10 units of thrombin (Sigma) were added per mg of protein and the resulting solution was left at either 18°C or 37°C for a recorded amount of time, in order to set up a time course experiment to ascertain the best temperature and time for complete cleavage. A solution containing wild type AcpS and thrombin was made and left at 18°C overnight. An affinity chromatography column (HisTrap HP 5 mL (Amersham)) was placed directly at the top of a benzamidine column (HiTrap Benzamidine FF, 1 mL (Amersham)) in the Akta prime (GE Healthcare). This system was equilibrated with equilibrating buffer (50 mM Tris pH 8, 0.5 M NaCl, 10 mM $MgCl_2$), then the solution was injected into the nickel column. Both columns were washed again using the same buffer. The untagged protein should be eluted immediately.

3.6.7 Dialysis

3,500 Da dialysis membranes (Medicell International Ltd) were washed in deionised water, then in dialysis buffer (Table 3.6). After sealing one end with a clip, the pooled protein fractions were inserted. The other end was then sealed and the membrane transferred into 1 L of the dialysis buffer at 4°C, for 1 h, under gentle agitation. The membrane was then placed in 1 L of fresh buffer, which was left at 4°C overnight. The next morning the buffer was replaced with 1 L of fresh dialysis buffer and left for a couple of hours at the same temperature.

Protein	Dialysis buffer
AcpS	50 mM Tris-HCl pH 8, 10% v/v glycerol, 10 mM $MgCl_2$

Table 3.6 Dialysis buffer used.

3.6.8 Desalting

The HiTrap Desalting 5 mL (Amersham) column was initially equilibrated with the appropriate desalting buffer (Table 3.7) at a flow rate of 1 mL/min. Then, up to 1.5 mL of protein solution was loaded and 0.3 mL fractions were collected.

Protein	Desalting buffer
ACP	50 mM Tris-HCl pH 8.5, 10 mM $MgCl_2$,10 mM DTT,
AcpS	50 mM Tris-HCl pH 8, 10% v/v glycerol 10 mM $MgCl_2$

Table 3.7 Desalting buffers used for each different protein.

3.6.9 Activation and purification of holo-ACP

For the phosphopantetheinylation of apo-ACP, 100 µL aliquots of reaction buffer (50 mM Tris, 10 mM $MgCl_2$, 5 mM DTT, pH 8.8) containing the reaction mix (100 µM ACP, 1 µM AcpS, 1 mM CoA) were left for 1 hour at 30°C in a water bath. Then samples were centrifuged and loaded into a gel filtration column (Superdex75 HiLoad 16/60 (Amersham) or Superdex75 10/300 GL (Amersham)), which was previously equilibrated with gel filtration buffer (50 mM Tris, 200 mM NaCl,10 mM $MgCl_2$, 5 mM DTT, pH 8.8).

3.7 Biochemical characterization

3.7.1 SDS-PAGE gels

SDS-PAGE gels were used to check the presence of protein after its expression and throughout the purification. Polyacrylamide gels are made using the following reagents (for 12 gels):

- for stacking gel (4% polyacrylamide): 15 mL deionised water, 3.75 mL 30% acrylamide (Severn biotech), 6.25 mL stacking gel buffer (30.3 g Tris, 10% SDS, pH 6.8). Then: 15µL Temed (Bio-Rad), 250 µL 10% APMS;
- for resolving gel (15% polyacrylamide): 17.5 mL deionised water, 35 mL 30% acrylamide, 17.5 mL resolving gel buffer (184.8 g Tris, 10% SDS, pH 6.8). Then: 45µL Temed, 700 µL 10% APMS. 10% or 8% stacking gels were made varying the volume of polyacrylamide added.

The protein samples were mixed with an equal amount of sample application buffer (30 mL glycerol, 50 mL 10% SDS, 10 mL 0.5 M EDTA, 5 mL β-mercaptoethanol, 0.1% Bromophenol blu, 0.6 g Tris, pH 7, for a final volume of 100 mL). This mixture was then heated to 100°C for about 10 minutes. 15 μL of this mix was usually loaded into each well, along with 5 μL low molecular weight marker (Sigma). Running buffer (25 mM Tris-HCl, 0.192 M glycine, 10% SDS) was used to make the gel run at 180-200V for about 1 hour, or until the visible blue dye reached the bottom of the gel. The gel was then removed and placed in Coomassie blue stain (500 mL stain solution contains 125 mL ethanol, 50 mL acetic acid, 1.25 g Coomassie brilliant blue and deionised water) on a rocker for about 10 minutes. After removing the stain solution, the gel was transferred into a destain solution (20% ethanol, 5% acetic acid) until clear bands were visible.

3.7.2 Protein concentration

The protein concentration was calculated using the Lambert-Beer equation:

$$A = \varepsilon c l$$

where A is the absorbance at 280 nm (measured in a Perkin-elmer lambda 25 Uv/vis spectrophotometer), ε the extinction coefficient for the protein (calculated using the software ProtParam [134]), c the concentration of protein in mg/mL and l the pathlength (1 cm). Protein was concentrated to the suitable value using a concentrator of the appropriate cut-off (3, 10 or 30 kDa Vivaspin 20, Sartorius), and then was frozen in liquid nitrogen and stored at -80°C until required.

3.7.3 Bradford assay

Bradford assay was used to determine protein concentration, especially when no tryptophan residues were found. Dilutions of BSA (1, 0.5, 0.25, 0.05 mg/mL) were used as standards. 20 μL of protein solution were diluted with 180 μL of Bradford reagent (Sigma-Aldrich), for a total volume of 200 μL. Each dilution was repeated at least twice. A Microtitre 96-well plate (Sterilin) was used and the absorbance at 595 nm measured with a Versa Max microplate reader (Molecular Devices).

3.7.4 Mass spectroscopy

Phosphopantetheinylation of apo-ACP was checked by NMR by the proteomic service at the Department of Biochemistry, Bristol. 10 μL protein samples were checked in a MALDI-MS Voyager-DE STR Mass Spectrometer (Applied Biosystems) to give a mass fingerprint. The purity of the cofactors used in the crystallization of AcpS was kindly checked by Chris Arthur, Department of Chemistry, using a positive ion mode on a QStar XL mass spectrometer (Applied Biosystems). Samples were delivered using a Nanomate nano-electrospray source (Advion) at a voltage of 1.3 kV and a gas pressure of 0.3 Psi.

3.8 X-ray crystallography

3.8.1 Crystallization trials

The hanging drop technique was used. The cofactor, when needed (generally at a final concentration of 5 mM), was added to the protein solution (usually 5 or 10 mg/mL). The mixture was left at 4°C or 18°C for 30 minutes, then centrifuged at the same temperature for 30 minutes at the maximum speed. For 24 well plates (24 well VDX plates with sealant, Hampton Research), 0.5-2 μL drop of protein solution at the suitable concentration were mixed with the same volume of crystallization reagent on the coverslip, and placed in vapour equilibration with 400 μL of liquid reservoir. The corresponding well was sealed and left at 18°C in the crystallization room or at 4°C in the fridge. Alternatively, the Phoenix robot (Art Robbins Instruments) was also used to test a very small volume of protein. Up to 0.2 μL drops (0.1 μL + 0.1 μL) were obtained using 50 μL of precipitant solution and 96 well plates (MRC 96 well plates, Molecular Dimension LTD). If a protein complex was required, proteins were incubated together for 30 minutes at a ratio of 1:1. The crystallization kits used were: Crystal Screens 1 and 2, PEG/Ion Screen (Hampton Research); Clear Strategy Screen I and II; Structure Screen I and II 96 HT; PACT-*premiere* 96 HT; JCSG-*plus* 96 HT (MDL); Jena Screens 1-9. The formulations of the screens are shown in Appendix 3.

3.8.2 Cryoprotectant and loop mounting

All crystals were flash-frozen in liquid nitrogen before testing them on synchrotron source. Crystals were mounted on a nylon loop of the suitable dimension (Hampton

research), transferred for 5-10 seconds into a stabilising solution containing the cryoprotectant (generally the reservoir plus 10-30% glycerol or 10-30% PEG400), then flash-frozen in liquid nitrogen.

3.8.3 Data collection and processing

Data were collected at the synchrotron SRS, Daresbury (beamline 14.1 / Quantum 4 ADSC detector and 10.1 / MAR 225 CCD detector), DLS, Diamond (beamline IO2 and IO3 / ADSC Q315 CCD detectors) and ESRF, Grenoble (beamline BM14 / MAR 345 image plate detector). The diffraction patterns were displayed, indexed, integrated and scaled using HKL2000 [135].

3.8.4 Molecular replacement

In preparation of subsequent datasets of the same crystal form, the same set of R-free reflections was used. The molecular replacement was performed using Phaser [136, 137]. The search models used to solve the different structures are summarized in Table 3.8.

Protein	**Search model**	**PDB**
Apo-AcpS (2jca)	*B. subtilis* apo-AcpS polyserine model	1f7t
AcpS in complex with CoA (2jbz)	*S. coelicolor* apo-AcpS	2jca
AcpS in complex with Acetyl-CoA	*S. coelicolor* holo-AcpS	2jbz
AcpS H110A	*S. coelicolor* holo-AcpS	2jbz
AcpS D111A	*S. coelicolor* holo-AcpS	2jbz

Table 3.8 Search models used for the molecular replacement.

3.8.5 3-fold averaging and density modification for apo-AcpS

After Phaser [136, 137], an electron density map was obtained, but it was difficult to build into. Structure factors and phases were calculated using SFall [138], and appropriate weights for the Fourier coefficients were then calculated using SigmaA [139]. Attempts to automate model building with ARP/wARP [140], with the option 'automated model building starting from existing model', using the model obtained after Phaser, failed. So, DM [141] was used to 3-fold average to the

electron density map and to generate new experimental phases. DM was run with the solvent, histogram, averaging and ncycle auto options. ARP/wARP [140] was then used to automatically trace the chain, utilizing the 'automated model building starting from experimental phases' option with 10 (50 in total) building cycles. Further rounds of model building in Coot [142] and refinement with Refmac5 [133] (provided with the CCP4 suites of programs [143]) resulted in the final model.

3.8.6 Refinement and model building

Refinement of the structures was done using Refmac5 [133], while the manual model building was made using Coot [142]. After a first cycle of rigid body refinement, restrained refinements were run. The software was used with the default options, except that a Babinet scaling was used and the overall weight for B-factor restrains were set at 1 (overall wt), 3 (sigmas bonded), 5 (non bond), 6 (planar) and 10 (H atom). This was to allow greater deviation along sidechians where the terminal portion can be more variable than parts close to the main chain. The improvement of the quality of the model was followed by the R_{free} and R-factor values. The geometric quality of each structure was analyzed by Procheck [144], Molprobity [145] and using the interactive tools in Coot. In this thesis, two type of electron density map were routinely used: FWT with phases PHWT, an equivalent to the 2Fo-Fc where the exact weight is determined using SigmaA as Refmac output, and the difference electron density map DELFWT with phases DELPHWT, equivalent of Fo-Fc.

3.8.7 Generation of AcpS-ACP complex model

The NMR structures of *S. coelicolor* apo- and holo-ACP, kindly provided by Dr Matt Crump, Department of Chemistry, were used to study the possible interaction with AcpS. The structure closest to the geometric mean was chosen from the ensemble of NMR solutions. *B. subtilis* AcpS-ACP complex (PDB: 1f80) was used as starting model. A trimer of *S. coelicolor* apo- or holo-AcpS was superimposed to the *B. subtilis* AcpS using the SSM superpose tool in Coot [142], while apo- or holo-ACP was superimposed on *B. subtilis* ACP (D subunit) and the complex was visualized in Coot.

3.9 ITC

K_d for wild type and mutants AcpS and cofactors binding was determined by ITC. The protein and cofactor were first dialysed in ITC buffer (50 mM Tris-HCl pH 8). A protein concentration of 8-10 μM and a cofactor concentration of 120 μM were used. The ITC was carried out using a VP-ITC microcalorimeter (MicroCal). A single experiment consisted of an initial injection of 7.5 μL of cofactor solution into the cell containing the protein solution, followed by other 18 injections of 15 μL. The reaction was carried out at 25°C and each injection lasted for 30 seconds (15 for the first one) followed by 240 seconds of pause. For the determination of the weaker CoA binding site of the D111A AcpS, an initial concentration of 1.5 mM of CoA was used. 15 injections of 2.5 μL of cofactor over 5 seconds were followed by other 15 injections of 15 μL over 30 seconds. For each cofactor or protein, the experiment was repeated at least 3 times. A blank that covered the concentration range of the appropriate cofactor was run as a negative control in the absence of protein and was subtracted from experimental data. Alternatively, ΔH determined at saturation of binding could be subtracted as this was determined to be linear from control experiment.

RESULTS AND DISCUSSION

Chapter 4

The Acyl carrier protein Synthase (AcpS)

Streptomyces coelicolor Acyl carrier protein Synthase (AcpS) catalyzes the transfer of the phosphopantetheine chain from coenzyme A (CoA) to a conserved serine of the Acyl carrier Protein (ACP). This chain has a core role during both fatty acid and polyketide synthesis. It holds and delivers the growing chain to the different enzymes involved in the syntheses. *S. coelicolor* AcpS is capable of activating both fatty acid and polyketide ACP [74]. This chapter presents the structure solutions of the apo-AcpS and of the protein in complex with CoA. The comparison of this type I *S. coelicolor* AcpS with the type I fatty acid *B. subtilis* and *S. pneumoniae* AcpS, and with the type II human AcpS, are presented in order to highlight the common aspects and the differences that characterize this promiscuous enzyme.

4.1 Sequence analysis

pET-16b-AcpS DNA was kindly donated by Dr Matt Crump and Dr John Crosby, Department of Chemistry. The sequence of the protein is shown in Figure 4.1. The protein was cloned with a polyhistidine tag at the N-terminal position, with the additional presence of a thrombin recognition site for cleavage of the tag. The software ProtParam [134] was used to determine the molecular weight of 14.7 kDa, the pI of 6.75 and the extinction coefficient of 1.120 M^{-1} cm^{-1}. No cysteines were present, so no reducing agents were used during the purification protocol. Finally, the presence of three tryptophanes and three phenylalanines guaranteed the UV absorbance at 280 nm. The poly-histidine tag was not removed during AcpS purification.

MGSSHHHHHHSSG **LVPRGS** H MSIIGVGIDVAEVERFGAALERTPALAG
RLFLESELLLPGGERRGVASLAARFAAKEALAKALGAPAGLLWTDAEV
WVEAGGRPRLRVTGTVAARAAELGVASWHVSLSHDAGIASAVVIAEG

Figure 4.1 The sequence of the *S. coelicolor* his_6-AcpS. The polyhistidine tag and the sequence preceding the AcpS gene are underlined. The LVPRGS thrombin cleavage site is shown in bold.

4.2 Expression of wild type AcpS

Before the expression, pET15b-AcpS DNA was transformed into *E. coli* BL21 (DE3) (chapter 3.4.5). The success of the transformation was checked by extracting plasmid from transformants (chapter 3.4.2), subjecting the DNA to endonuclease restriction digest (chapter 3.4.3) and checking the size of the fractions on an agarose gel.

AcpS was expressed by induction with 1 mM IPTG at 37°C, and the expression continued for 4-5 hours at 30°C (chapter 3.5). AcpS was expressed very well in both 0.5 L and 1 L culture. Some protein can be seen in the insoluble fraction, but the amount of soluble protein obtained after sonication (Figure 4.2) was still very satisfactory for purification purposes.

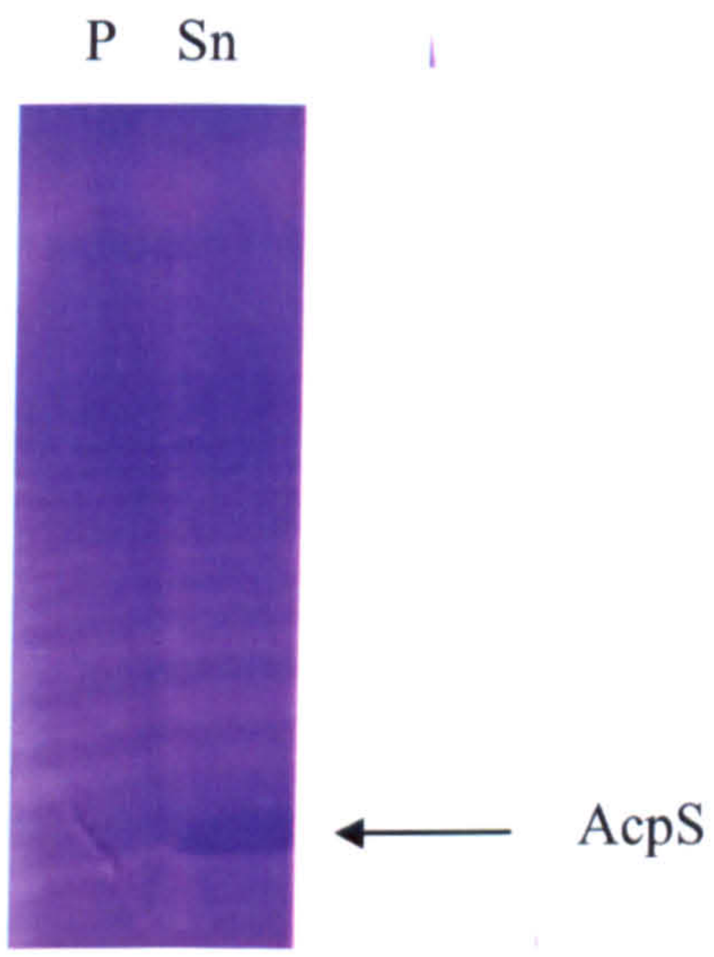


Figure 4.2 AcpS expression. P: pellet after sonication; Sn: supernatant after sonication.

4.3 Purification of wild type AcpS

The supernatant after sonication was loaded into the affinity column (chapter 3.6.3). AcpS was eluted from the column as a single and well defined peak, at a concentration of imidazole of about 120 mM (Figure 4.3 A). The SDS-PAGE gel (Figure 4.3 B) also showed that the fractions after the affinity chromatography were very pure. To increase protein purity, as well as for analytical purposes, aliquots from this peak were collected, concentrated to 1 mL and loaded into the Superdex 75 gel filtration column (chapter 3.6.5).

As shown in the chromatogram and in the SDS-PAGE gel (Figure 4.4), AcpS was eluted again as a very well-defined single peak. Aliquots from this peak were dialysed or desalted in the desalting buffer (chapter 3.6.7-3.6.8) and then stored at -80°C. The protein concentration was determined by UV absorbance and Bradford assay. About 10 mg of pure AcpS were obtained from 0.5 L of bacterial culture.

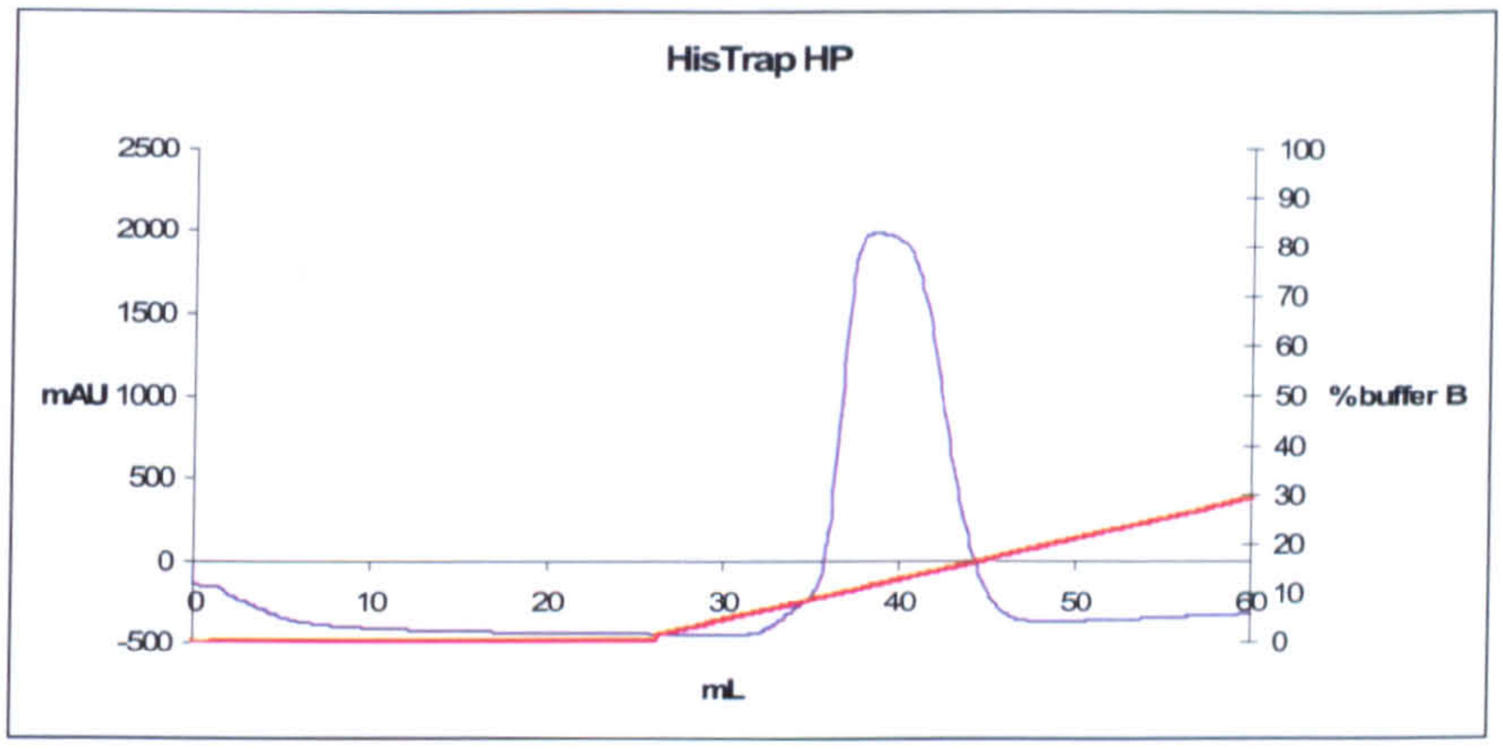
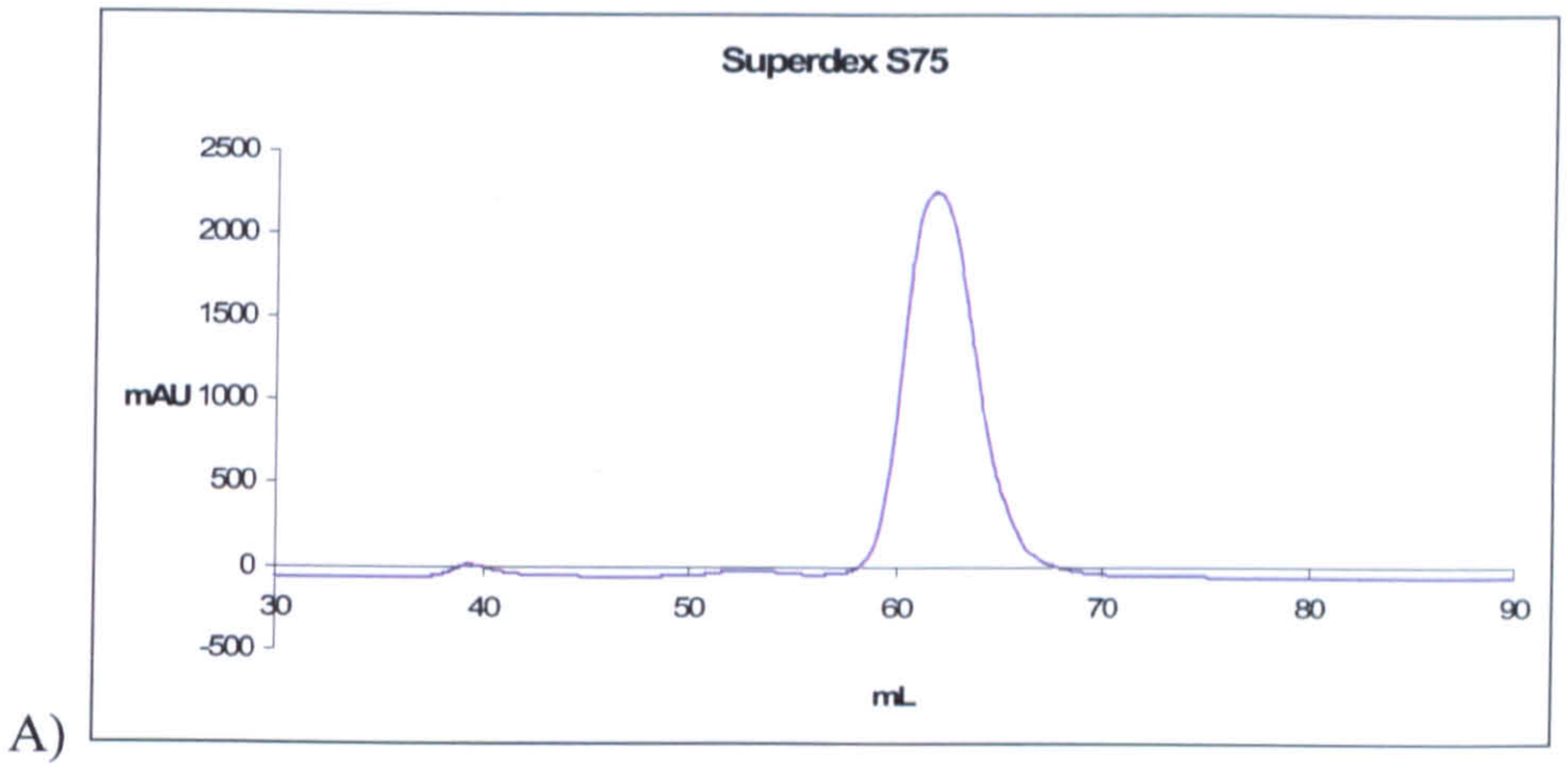


Figure 4.3 AcpS affinity chromatography. A) Chromatogram from the HisTrap column. AcpS is eluted as a well defined single peak. UV absorbance is shown in blue, the gradient of the elution buffer in red. B) SDS-PAGE gel showing AcpS fractions pooled from the peak of the chromatogram. MW: molecular weights; W: aliquots after column washes with equilibrating buffer; F: AcpS fractions.



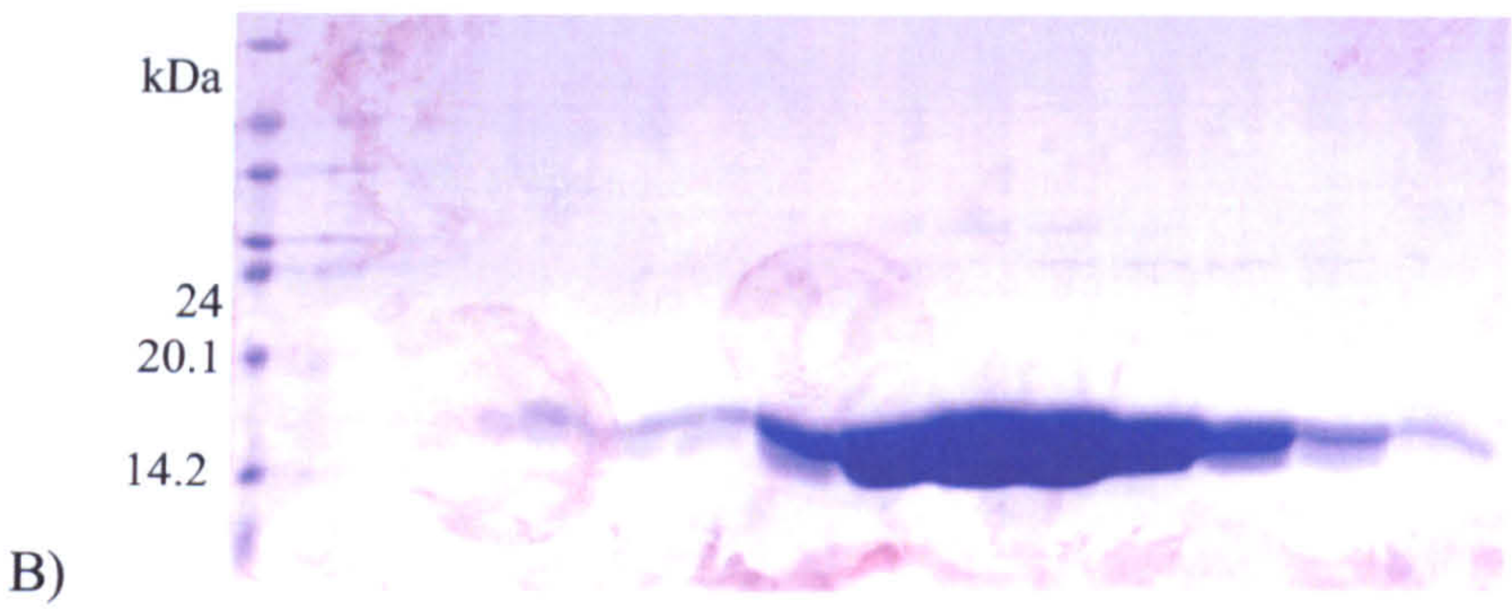


Figure 4.4 AcpS gel filtration chromatography. A) Chromatogram from the gel filtration column. AcpS is eluted as a single peak. UV absorbance is shown in blue. B) SDS-PAGE gel showing AcpS fractions pooled from the peak of the chromatogram eluted between 58 and 69 mL (1 mL fractions). MW: molecular weights; F: AcpS fractions.

4.4 AcpS crystallization, data collection, processing and refinement

4.4.1 Apo-AcpS data collection and processing

The apo-AcpS was purified and crystallized by Dr Chris Arthur, Department of Chemistry, using the hanging drop vapour diffusion method. A crystal obtained at 18°C, using 10 mg/mL protein concentration, 0.2 M NH_3SO_4, 0.1 M NaCacodylate pH 6.5, 30% PEG 8000 as precipitant (Figure 4.5), was donated and tested at the beamline 14.1 at the Synchrotron Radiation Source (SRS), Daresbury. The crystal, about 1.2×0.1×0.1 mm, was transferred into 0.2 M ammonium sulphate, 0.1 M NaCacodylate pH 6.5, 30% PEG 8000, 10% glycerol prior to flash freezing in liquid nitrogen. A nice protein-like diffraction pattern was obtained (Figure 4.5).

This crystal belonged to the trigonal space group $P3_121$ with a=b=75.9 Å and c=108.3 Å, and diffracted to a resolution of approximately 2 Å. The data processing was done using the HKL2000 [135].

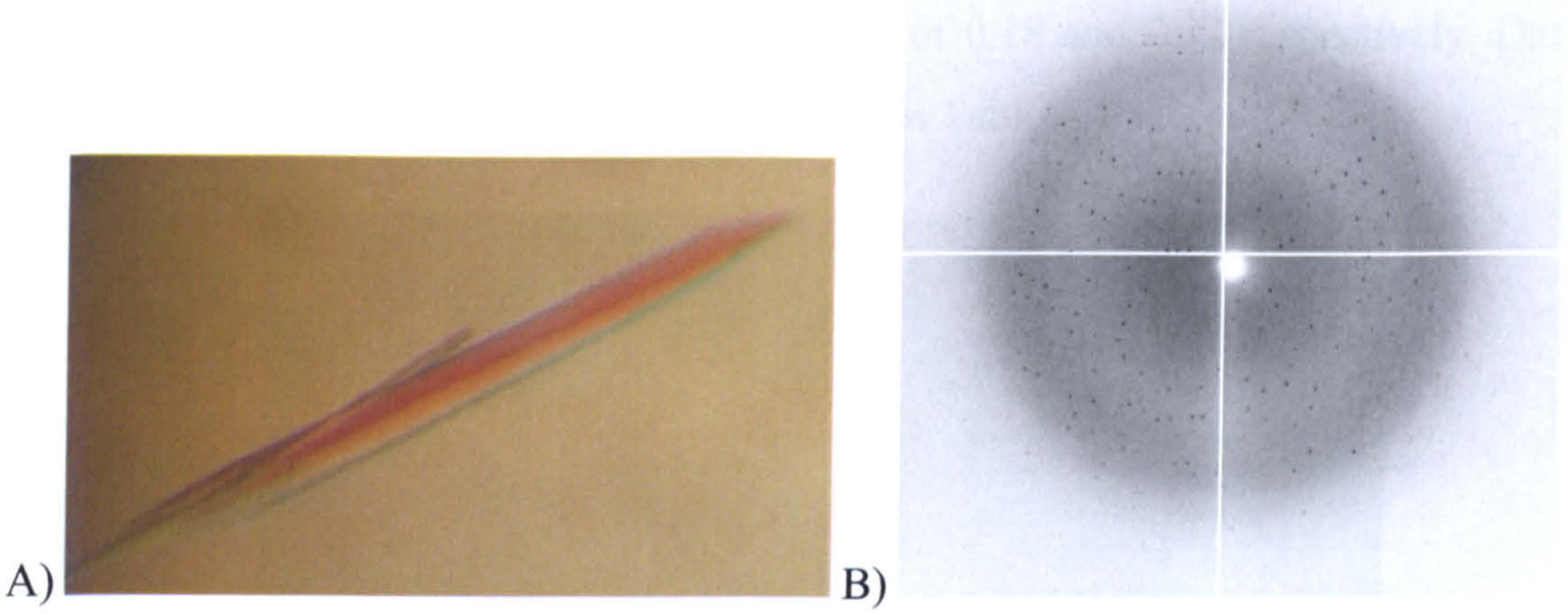


Figure 4.5 Apo-AcpS crystal and related diffraction pattern. A) Crystals obtained in 0.2 M ammonium sulphate, 0.1 M NaCacodylate pH 6.5, 30% PEG 8K. B) Example of diffraction image.

4.4.2 Molecular replacement and refinement

Calculation of possible solvent content suggested that there were 3 molecules per asymmetric unit. The structure was solved by molecular replacement using Phaser [136, 137] and the *B. subtilis* FAS AcpS protein (PDB: 1f7t [86]) as search model (chapter 3.8.4). ARP/wARP [140] was not able to build the structure automatically into the initial electron density map using the existing Phaser model (chapter 3.8.5). So, 3-fold averaging was used to obtain better phases. After 3-fold averaging with DM [141] (chapter 3.8.5), the electron density map was much improved (Figure 4.6). The NCS averaging correlation between related density regions was enhanced of about 30% (from 57% to 87% and from 54% to 87% for the correlation between monomers A and B and monomers A and C respectively). ARP/wARP was then run using the structure factors and phases obtained with DM (chapter 3.8.5), and automatically built 337 out of 429 residues (including the histidine tags) into this averaged map. Further rounds of model building in Coot [142] and refinement with Refmac5 [133] resulted in the final model. The apo-AcpS model was deposited in the PDB as 2jca, and contains 123 residues per each molecule, 3 sulphate ions, 1

calcium, 1 glycerol and 199 water molecules. The histidine tag resulted very disordered, and it was not possible to build any model for it. 3 protein subunits (1 complete trimer) are found in the asymmetric unit. The electron density was weak and not well defined for residues 65-68 in the B subunit, and for residues 22-24 in the C subunit, due to the high mobility of these loops. The refinement led to a structure model with R-factor and R_{free} values of 0.18 and 0.22 respectively. Data processing and refinement statistics are shown in Table 4.1

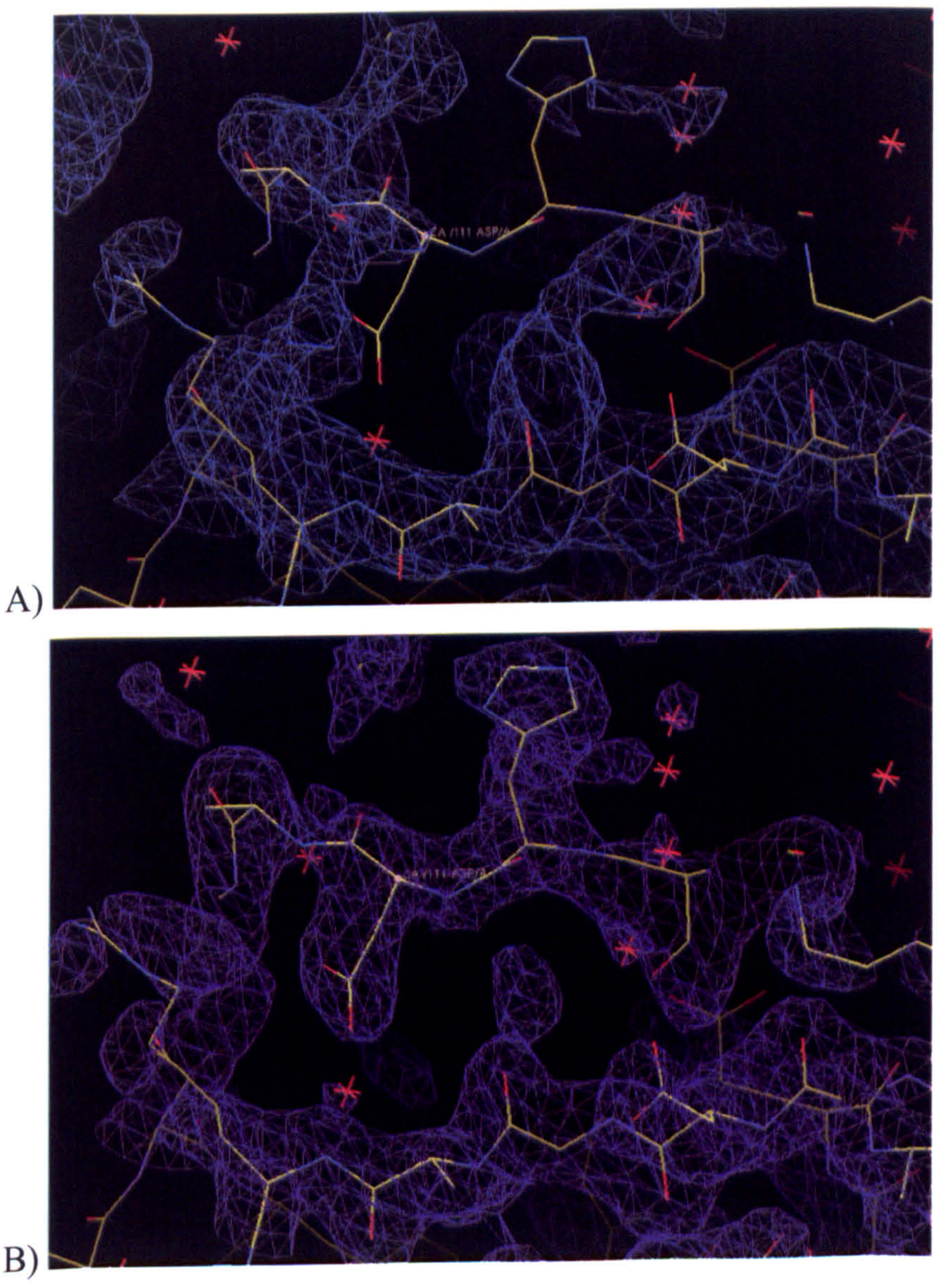


Figure 4.6 Map improvement after 3-fold averaging. A) 2Fo-Fc density map (FWT, PHWT) after molecular replacement solution contoured at 1σ with the final model shown as bonds. B) 2Fo-Fc electron density (FDM, PHIDM) obtained using the structure factors and phases from DM contoured at 1σ, with the final model shown as bonds.

	Apo ACPS [2jca]
Wavelength (Å)	1.488
Resolution (Å)	50-1.98 (2.05-1.98)
Space group	P $3_1$21 a=b=75.9 Å, c=108.3 Å
Completeness (%)	99.3 (95.7)
N unique reflections	25759
Redundancy	6.4 (3)
I/σI	22.2 (2.0)
R merge	0.08 (0.38)
Mosaicity (°)	0.37
R_{free}	0.22
R-factor	0.18
RMS bond (Å)	0.018
RMS angle (°)	1.690

Table 4.1 Data processing and refinement statistics for the apo-AcpS. Data corresponding to the highest resolution shell are shown between brackets.

4.4.3 Geometry and B-factor analysis

MolProbity [145, 146] was used to check the geometry and the conformation of the model. The Ramachandran plot is presented in Figure 4.7. 98.1 % of the residues are in the most favoured region, while Asp111 of subunits A and B, and Ala67 and Leu70 of subunit C, are in the allowed region. 3 outliers were identified: Thr23, Pro24 and Ala25 of subunit C. These residues belong to a mobile area, with a very high B factor, and their model building was quite difficult to complete. Met1, Leu49 and Leu108 of subunit A, Ser2, Arg43 and Leu108 of subunit B, and Leu70 and Val101 of subunit C were recognized to have unusual rotamers. In most of the cases, this is probably due to the fact that the electron density of the side chains is not very well defined. An analysis of the average B-factor of the main chains was also performed with Baverage [143] (Figure 4.8). The average B-factor is 32 $Å^2$, 36 $Å^2$ and 42 $Å^2$ for monomer A, B and C. This analysis shows that the average B-factor of the areas that will be identified as mobile (chapter 4.6.3) is higher than the

one of the neighbouring regions. The difference of the B factor values between each monomer is due to different crystal contacts.

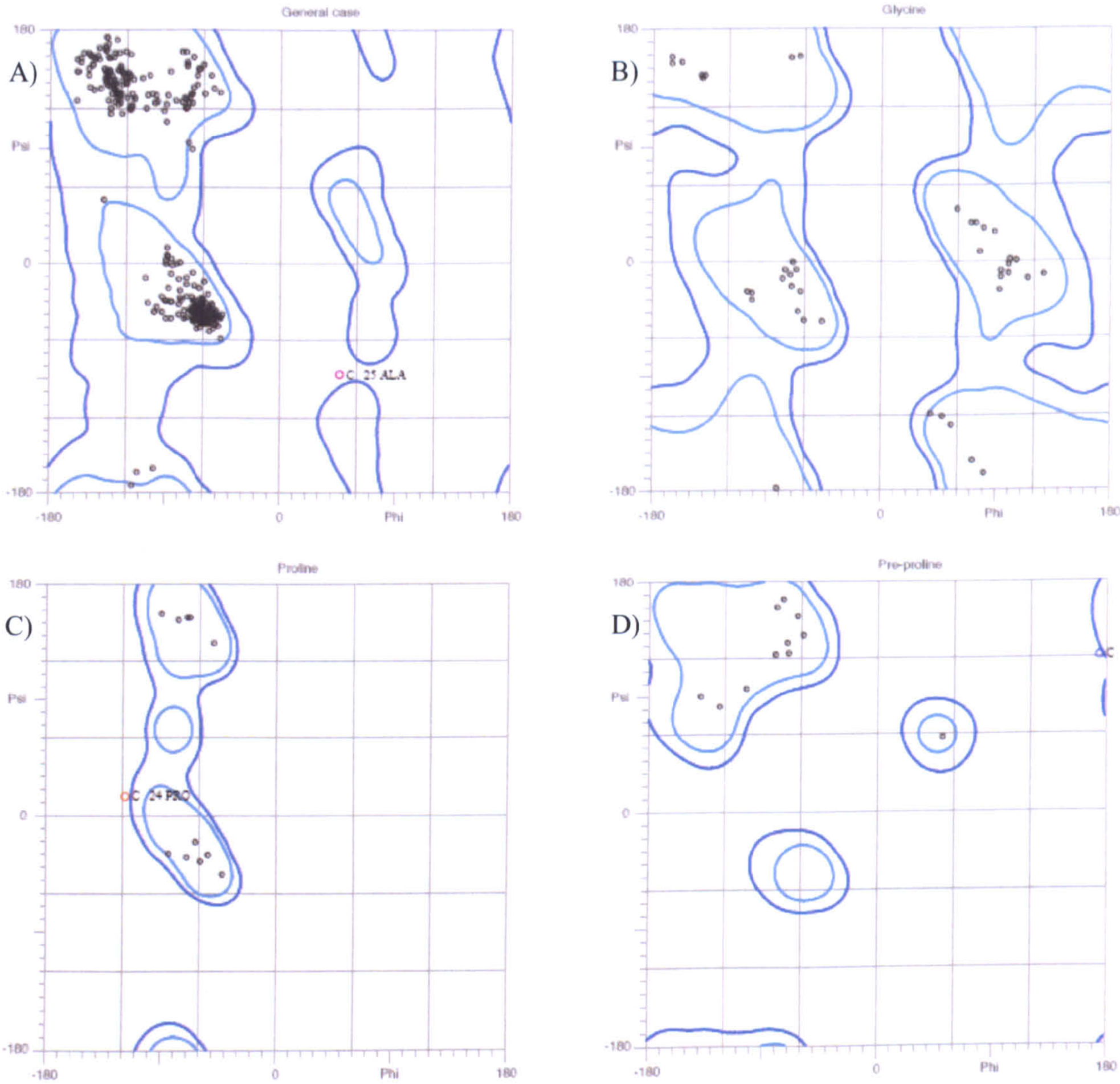


Figure 4.7 Ramachandran plot related to the apo-AcpS. This analysis was performed by MolProbity. A) general case; B) glycine; C) proline; D) pre-proline.

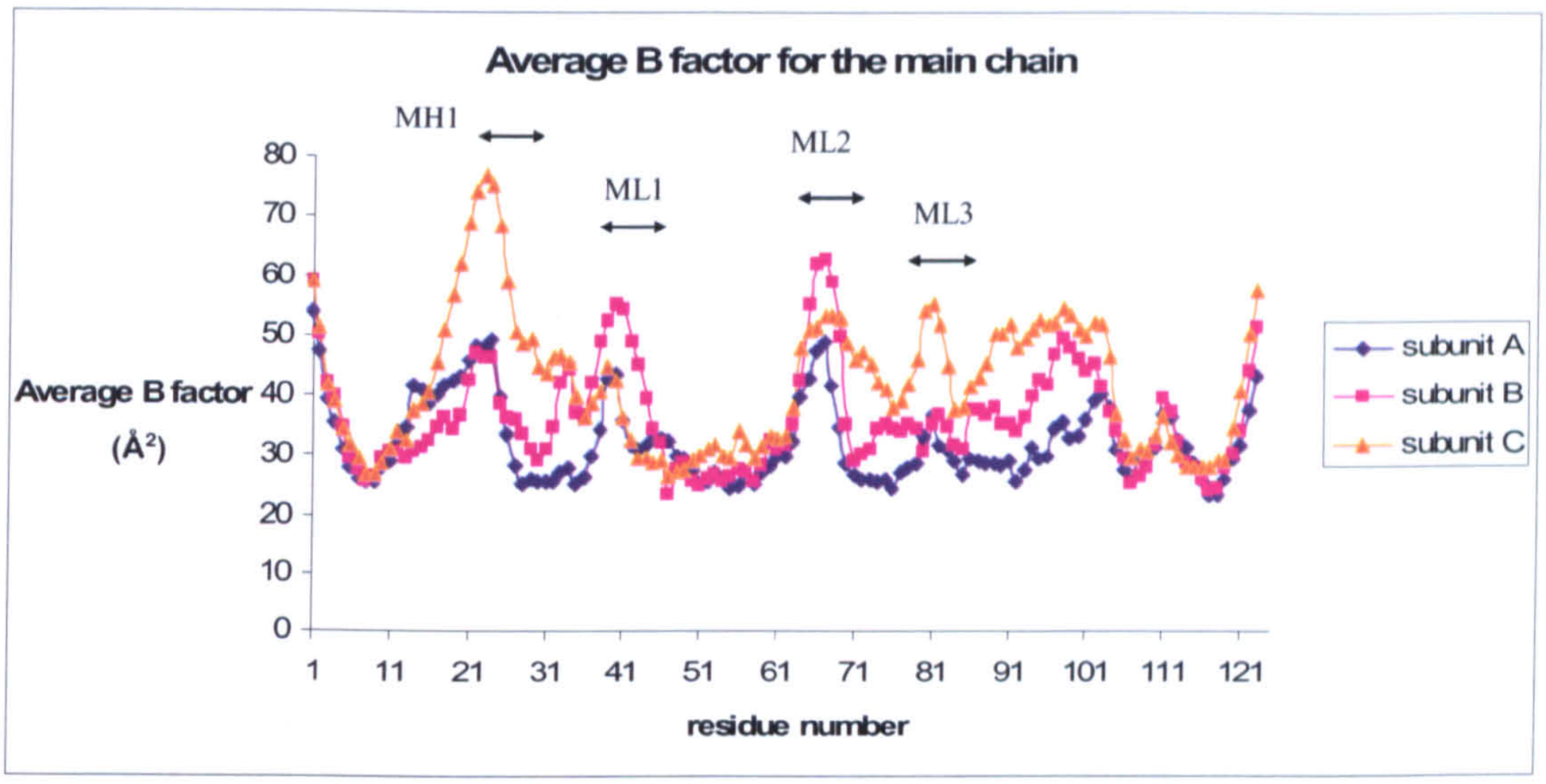


Figure 4.8 Apo-AcpS B-factor analysis. Average B-factor values for the main chain of the three subunits of the apo-AcpS. The B-factor values reflect the mobility of the mobile loops.

4.5 AcpS in complex with CoA

4.5.1 Holo-AcpS crystallization and data collection

AcpS at the concentration of 10 mg/mL was crystallized in the presence of 5 mM CoA at 18˚C (chapter 3.8.1). Very nice and quite big crystals were obtained when the Peg/Ion and the Clear Strategy Screen I and II at pH 6.5 were used. The most regular and big crystals were obtained at 18ºC in the conditions shown in Table 4.2.

Salt	Buffer	PEG
0.2 M lithium sulphate	0.1 M NaCacodylate pH 6.5	25% PEG 2K MME
0.2 M lithium sulphate	0.1 M NaCacodylate pH 6.5	15% PEG 4K
0.2 M potassium thiocyanate	0.1 M NaCacodylate pH 6.5	15% PEG 4K
0.2 M potassium thiocyanate	0.1 M NaCacodylate pH 6.5	10% PEG 8K+ 10% PEG 1K
0.2 M potassium bromide	0.1 M, NaCacodylate pH 6.5	25% PEG 2K MME
0.2 M potassium bromide	0.1 M NaCacodylate pH 6.5	15%PEG 4K
0.2 M potassium bromide	0.1 M NaCacodylate pH 6.5	10% PEG 8K+ 10% PEG 1K
0.2 M Lithium Nitrate	none	20% w/v Polyethylene Glycol 3350
0.2 M Ammonium Sulfate	none	20% w/v Polyethylene Glycol 3350

Table 4.2 Holo-AcpS crystallization conditions.

Optimization plates were prepared varying the salt (0.1, 0.2 and 0.3 M) and/or the PEG concentration (variations in steps of 1% for PEG 4K , between 13-18% w/v, and in steps of 2% for PEG 2K MME, between 19-27% w/v). Very nice triangular crystals were obtained in many different conditions, with an average size of about 0.7×0.6×0.3 mm (Figure 4.9). Cryosolutions consisting of the appropriate reservoir with the presence of 10 or 20% v/v glycerol were then prepared and the crystals were tested at beamline 10.1, Daresbury synchrotron. The best diffraction pattern was obtained using crystals from 0.2 M potassium thiocyanate, 15% PEG 4K, 0.1 M NaCacodylate pH 6.5. The best crystal, in a cryosolution containing 0.2 M potassium thiocyanate, 15% PEG 4K, 0.1 M NaCacodylate pH 6.5, 10% v/v glycerol, diffracted to approximately 1.6 Å. Data were collected at beamline 10.1 at the SRS synchrotron, Daresbury. Crystals belonged to the cubic $P2_13$ space group with a=b=c=73.1 Å. The data processing was carried out using the HKL2000 [135].

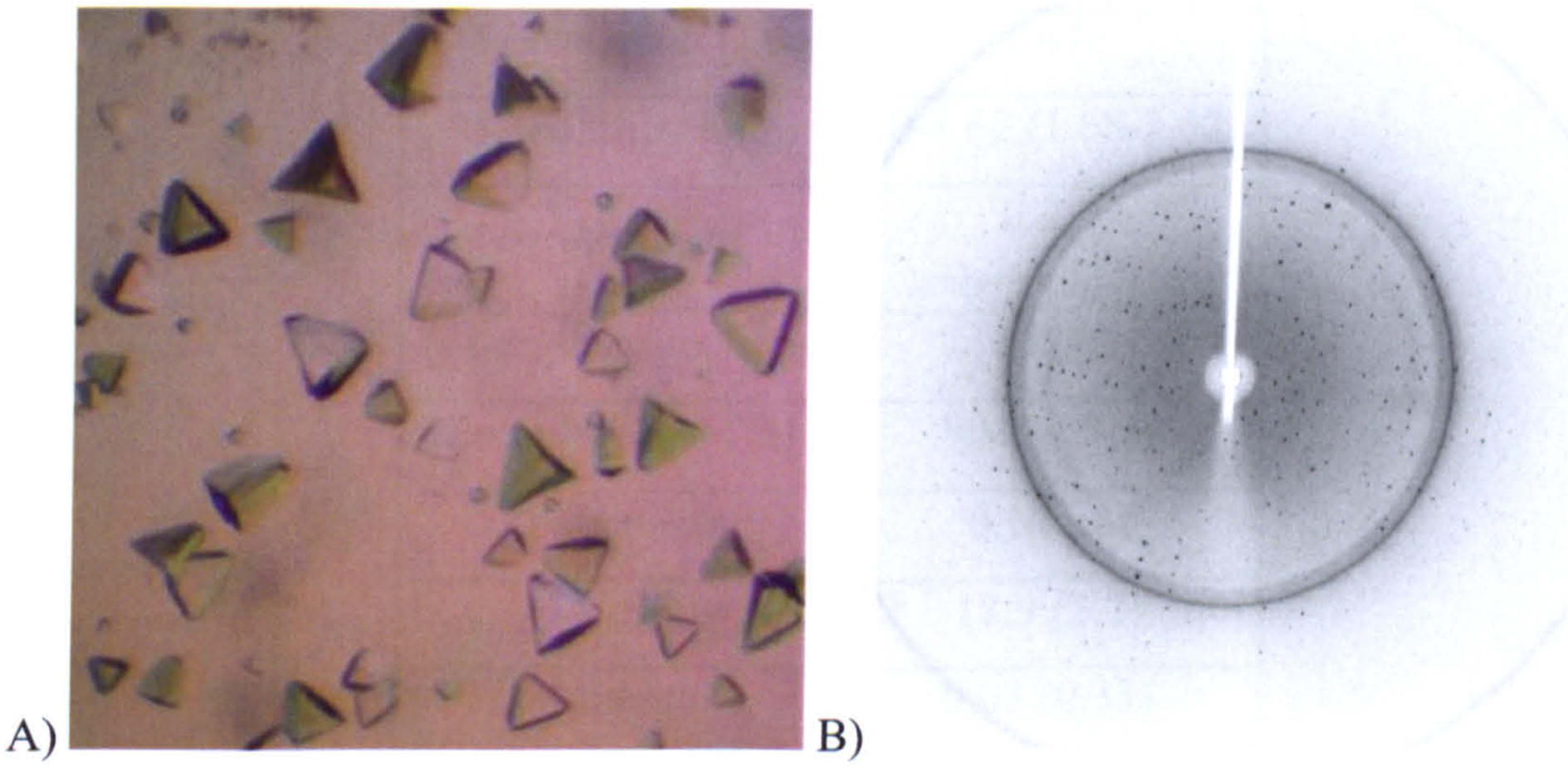


Figure 4.9 Holo-AcpS crystals and example of diffraction image. A) Crystals obtained in 0.2 M potassium thiocyanate, 15% PEG 4K, 0.1 M NaCacodylate pH 6.5. B) An example of diffraction image of the best crystal.

4.5.2 Molecular replacement and refinement

One protein subunit was found per asymmetric unit, with the trimer axis therefore lying on a crystallographic 3-fold axis. The apo-AcpS structure was used as search model for molecular replacement. Structure refinement was performed using Coot [142] and Refmac5 [133]. The holo-AcpS model was deposited in the PDB as 2jbz. It contains residues 0-123 of an AcpS monomer, a molecule of CoA, one magnesium ion, one potassium ion and 183 water molecules. Each molecule is organized in a trimeric oligomerization state. Whereas in the apo structure the entire his-tag was disordered, in the holo structure it was possible to model the histidine (residue 0) preceding the portion containing the His_6-tag, with further density likely to be the N-terminal portion which was difficult to model. The refinement led to a structure model with R-factor and R_{free} values of 0.19 and 0.21 respectively. Data processing and refinement statistics are summarized in Table 4.3.

	Holo-ACPS [2jbz]
Wavelength (λ)	1.478
Resolution (Å)	50-1.62 (1.68-1.62)
Space group	P $2_1$3 a=b=c=73.1 Å
N unique reflections	16798
Completeness (%)	90.7 (69.3)
Redundancy	6.1 (1.8)
I/σI	18.9 (2.2)
R merge	0.08 (0.33)
Mosaicity (°)	0.5
R_{free}	0.21
R-factor	0.19
RMS bond (Å)	0.011
RMS angle (°)	1.417

Table 4.3 Data processing and refinement statistics for AcpS in complex with CoA. Data related to the highest resolution shell are reported between brackets.

4.5.3 Geometry and B-factor analysis

The geometry and the conformation were checked using MolProbity [145, 146]. The Ramachandran plot is presented in Figure 4.10. 99.2 % of the residues are in the most favoured region, while Ser2 is in the allowed region. Leu59 and Arg83 were recognized to have unusual rotamers (0.2% and 0.1% respectively). An analysis of the average B-factor of the main chains was also performed with Baverage [143] (Figure 4.11). The average B-factor is 23 $Å^2$. From this analysis, it is also evident that the average B-factor of the areas that will be identified as mobile (chapter 4.6.3) is higher than the one of the neighbouring regions.

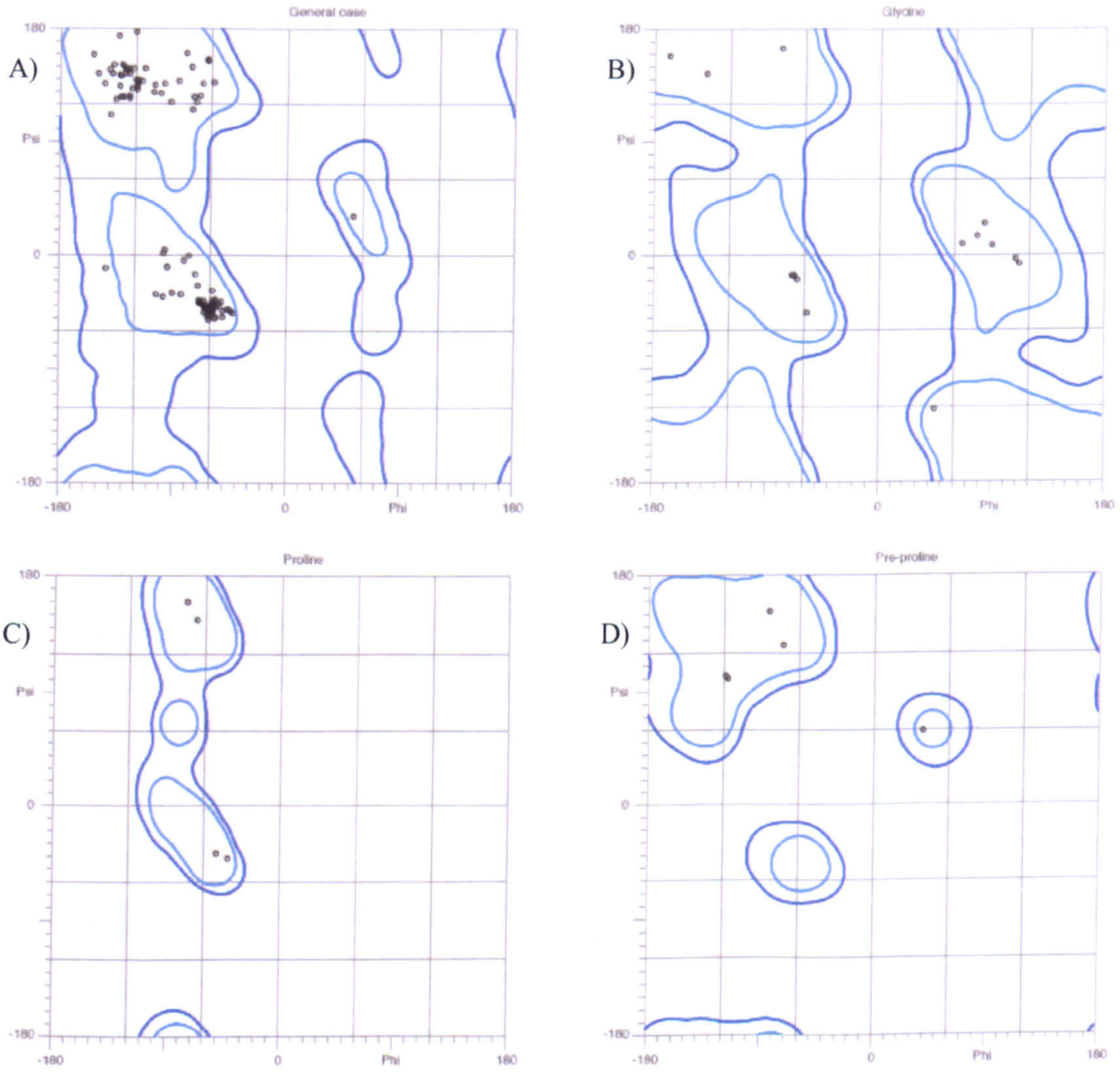


Figure 4.10 Ramachandran plot related to holo-AcpS. All residues are in the favoured regions with the exception of Ser2, that are located in the allowed region. A) general case; B) glycine; C) proline; D) pre-proline.

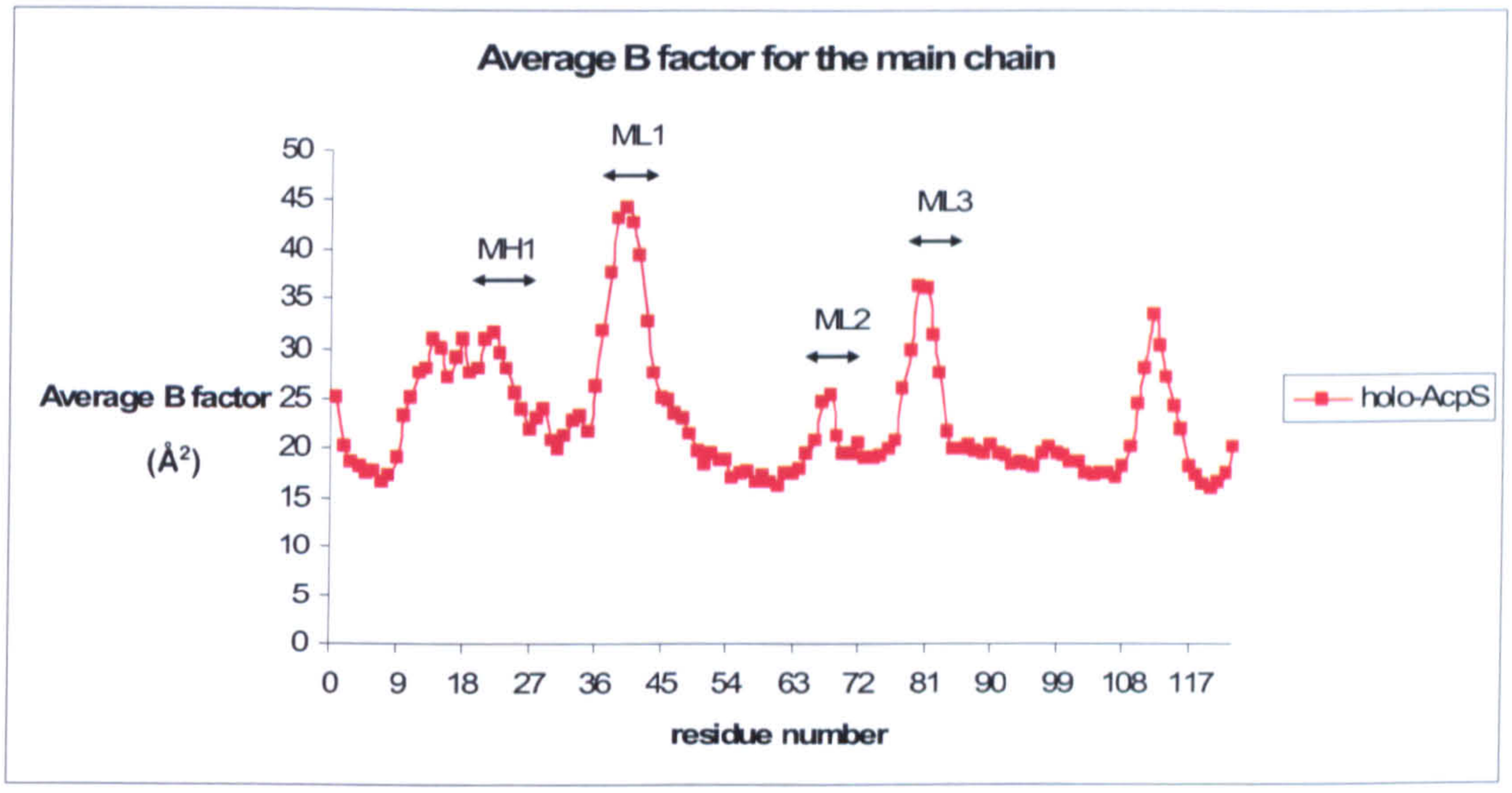


Figure 4.11 Holo-AcpS B-factor analysis. Average B-factor values for the main chain of AcpS in complex with CoA. The B-factor values reflect the mobility of the mobile loops.

4.6 Structural analysis of the Acyl carrier protein Synthase

In the following section, the structure of the holo-AcpS is introduced first, then the structure of the apo-AcpS is described and compared to the holo-AcpS.

4.6.1 AcpS fold and oligomerization

Holo-AcpS is a trimer, with the three fold axis of the trimer coinciding with a crystallographic 3-fold axis so that there is one subunit in the asymmetric unit. Each subunit is made up from 4 α helices and 5 β strands (Figure 4.12). The core of the subunit is helix α3, which runs between residues 45 and 63, and packs against a 3-stranded antiparallel β sheet on one side, with the other α helices decorating the other side of the subunit. This β sheet has topology β1, β4 and β5. To form the biological unit, a trimer with one active site at each interface between 2 subunits, the 3-stranded antiparallel β sheets come together around a 3-fold axis to make a 9-stranded β barrel with parallel interactions between the β sheets.

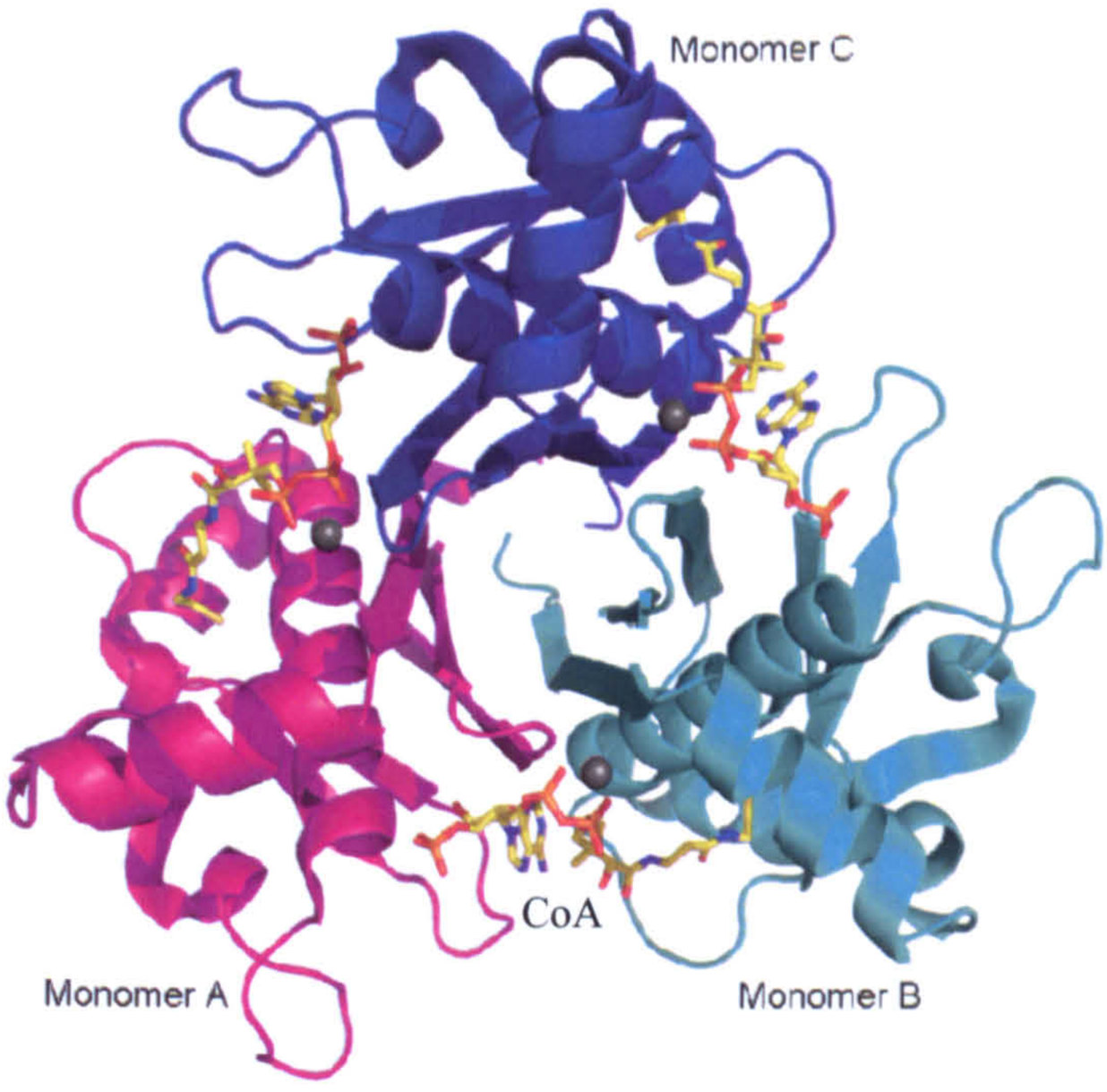


Figure 4.12 Structure of *S. coelicolor* AcpS. Diagram of the holo-AcpS biological trimer with the protein shown in cartoon representation, CoA as sticks and coloured according to atom type and magnesium shown as a grey sphere.

4.6.2 Active site

The active site and CoA binding site lie at the interface between 2 monomers (Figure 4.14). The adenine ring interacts with one monomer, while the phosphopantetheine arm binds to the adjacent monomer. The electron density of the cofactor is shown in Figure 4.13.

4.6.2.1 CoA binding site

In the following description the numbering describes the active site at the interface between A and B, but there are 3 equivalent active sites in the whole biological trimer, one at each subunit interface. The adenine ring binds in a pocket that is formed by the end of helix α3 on monomer B, and the ML3 of the adjacent A molecule. The ribose is present in the 3' *endo*-conformation. The 3'-phosphate is held in place by interactions with residues HisA110, SerA48 and ArgA52 and through hydrogen bonds provided by waters. The α-phosphate is within hydrogen bonding distance with magnesium, AspB9, LysB61 and the main chain of HisA110, while the β-phosphate with AspB9 and 3 water molecules.

The phosphopantetheine arm lies in a hydrophobic cavity on B between the central helix α3 and the small helix α5. At the entrance of the cavity, GluB57 makes a hydrogen bond to a backbone nitrogen at the phosphopantetheine side chain, (PN8), and side chain carbonyl oxygens (PO5 and PO9) make water mediated hydrogen bonds with ProB66 and AlaB67 in ML2. The hydrophobic cavity is composed of the aliphatic part of the LysB56 side chain and by residues LeuB59, LeuB63, LeuB69, TrpB71, AlaB74, ValB92.

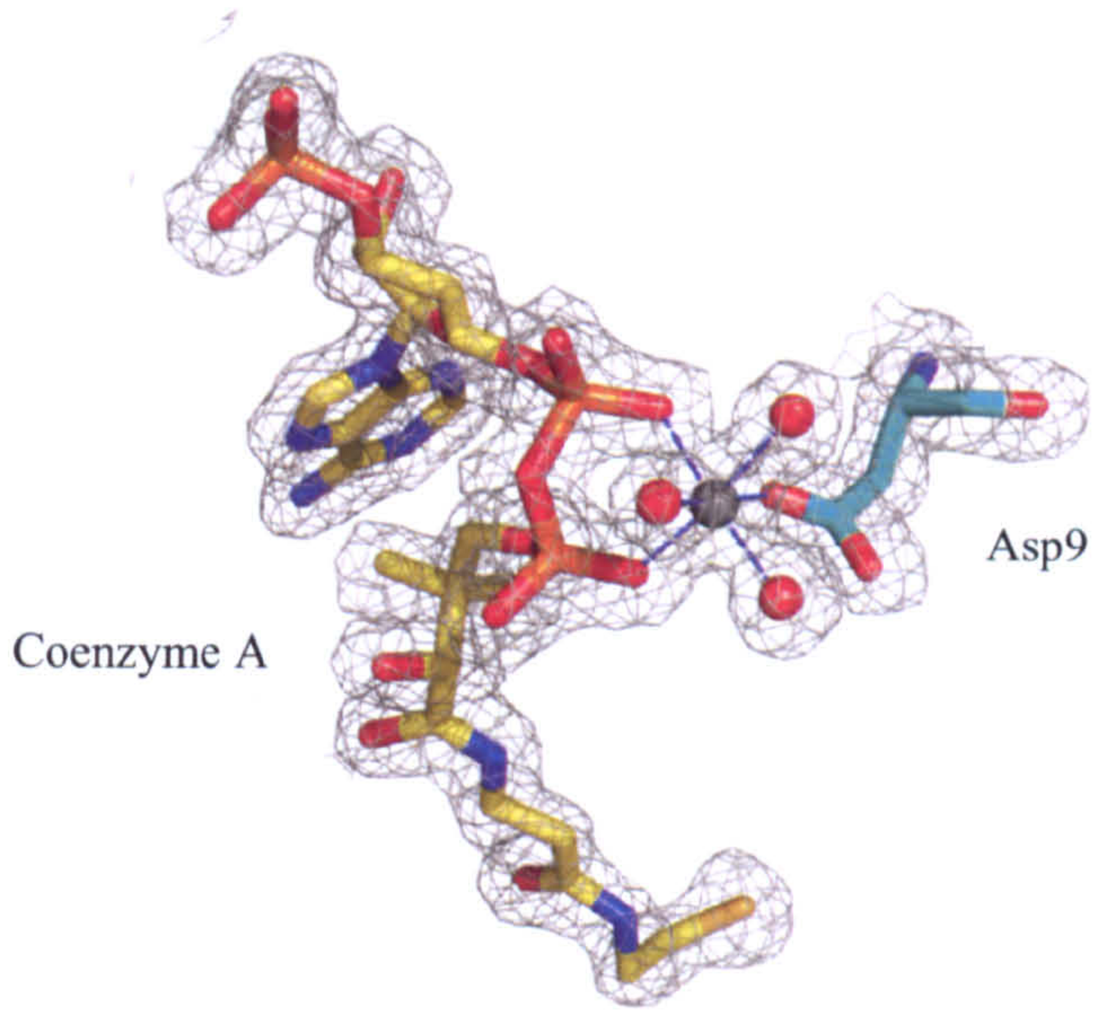


Figure 4.13 Electron density map (FWT, PHWT) at 1.2 σ of CoA and the magnesium coordination sphere. CoA in shown as sticks and coloured according to atom type and magnesium is shown as a grey sphere. The magnesium ion is coordinated to the α- and β-phosphate of CoA, to 3 water molecules (in red) and to an oxygen of the side chain of Asp9.

4.6.2.2 Magnesium binding site

Magnesium is essential for activity. This ion was present in the crystallization buffer and is clearly visible in the electron density in the presence of CoA. *S. coelicolor* holo-AcpS is the first type I AcpS structure where the essential magnesium ion is present in the active site. Its coordination sphere is composed of the oxygens of the alpha and beta-phosphates of CoA, the oxygen from the side chain of AspB9 and 3 waters molecules (Z12, Z16 and Z183) (Figure 4.13). One coordinating water (Z12) interacts with the oxygens of the beta-phosphate of CoA, AspB9, two water molecules (Z16 and Z183), and water Z175 from monomer A. The second water (Z16) is held in place by hydrogen bonds with the oxygens of the

alpha phosphate of CoA and of the backbone of ValB10, the side chain oxygen from SerA109, and 2 water molecules (Z12 and Z183). The third water (Z183) is within hydrogen bonding distance with the oxygen of the alpha phosphate, the side chain oxygen of AspA111, the other 2 coordinating waters, and water Z172 from monomer A.

4.6.3 Comparison between apo- and holo-AcpS

In the apo crystals, a complete trimer with 3 subunits is found in the asymmetric unit. The Cα atoms of the three subunits can be superimposed with a root mean square difference (RMSD) of 0.55 (A on B), 0.72 (A on C) and 1.05 Å (B on C). Three main flexible loops (ML1-3) are found between residues 37-44, 65-70 and 79-82 (Figure 4.14). There is also a small displacement in the final portion of helix α1, especially between residues 21-26 (MH1). The average B factor of these mobile portions is higher than in the other region. The Cα superimposition is improved to 0.29 (A on B), 0.39 (A on C) and 0.46Å (B on C) when these areas are omitted. The magnesium is not visible in the electron density in the absence of CoA, even if it is present in the crystallization buffer at the concentration of 10 mM. A sulphate ion, present in the mother liquor for the crystallisation, was found at the position where the 3'-phosphate of CoA is located in the holo-AcpS.

By superimposing a trimer of holo-AcpS generated using crystallographic symmetry on a trimer of apo-AcpS, some conformational changes can be identified. The RMSD after Cα superimposition of the A monomers, is 0.94 Å. The areas identified as mobile loops in the apo-protein are ordered in the holo model and are displaced compared to the apo. Residues in ML1 which are displaced by about 2 Å between the holo- and apo-structures, including ArgA44, are in the binding pocket for the 3'-phosphate. ML2 from the B subunit lines the phosphopantetheine binding pocket and ML3 stacks against the adenine base of the CoA. The loop that contains AspA111, which interacts with a magnesium coordinating water, is also displaced up to 6 Å. Moreover, by superimposing the 2 trimers, with a rigid superimposition of monomers A, it is possible to notice that helix α1 is displaced by 4-5 Å in chains B and C, clearly indicating a change in the trimer arrangement in the active holo-protein compared to the apo-protein, which results in an "opening up" at the top of the trimers, as displayed in Figure 4.15.

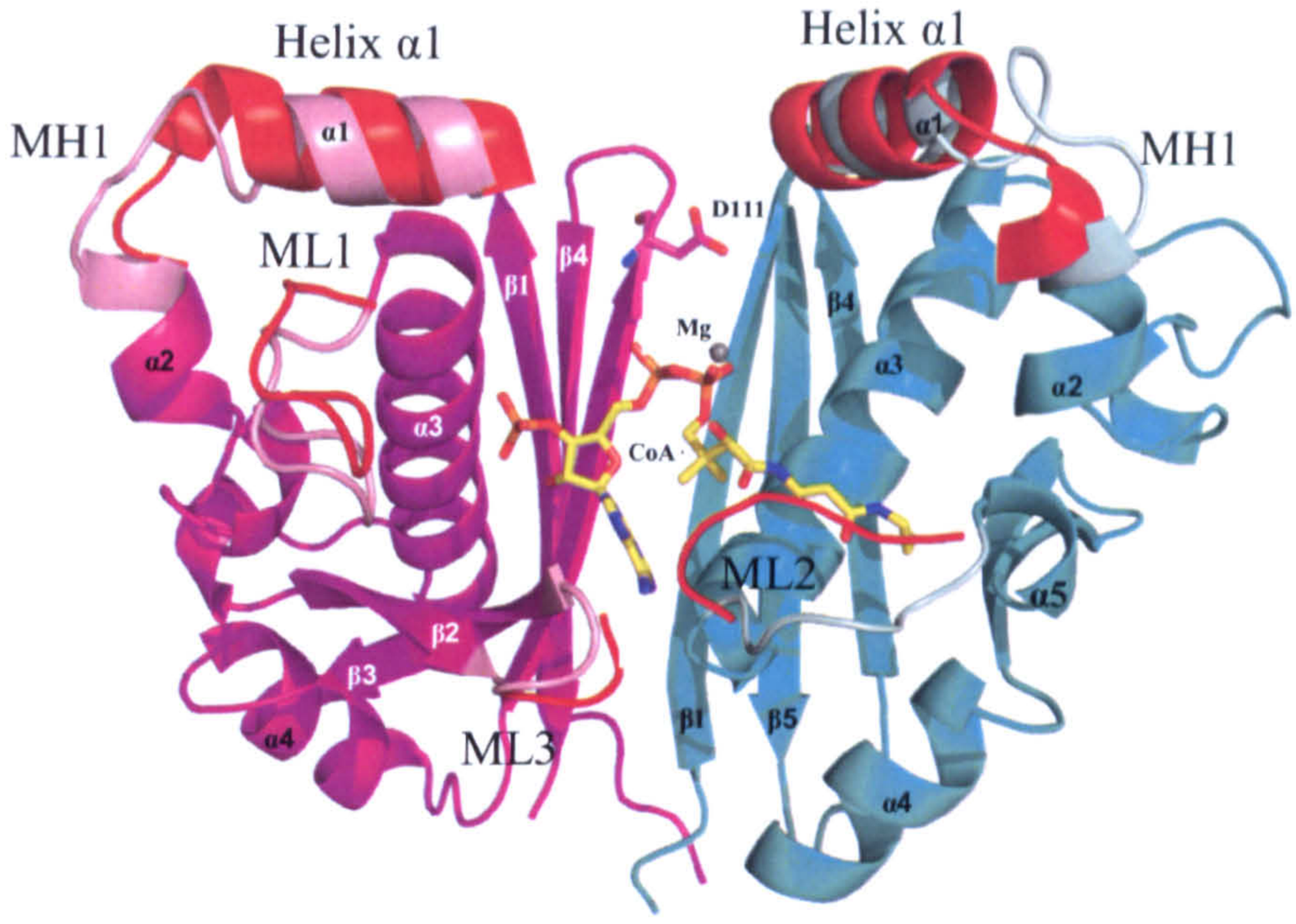


Figure 4.14 CoA lies at the interface between 2 monomers. Cartoon representation of holo-AcpS monomers A and B, coloured in magenta and cyan. CoA is shown as sticks and coloured according to atom type. Magnesium is shown as a grey sphere. Helix α1 and mobile loops in the apo-AcpS are colored in red. The equivalent positions in the holo protein are colored in pale magenta and cyan.

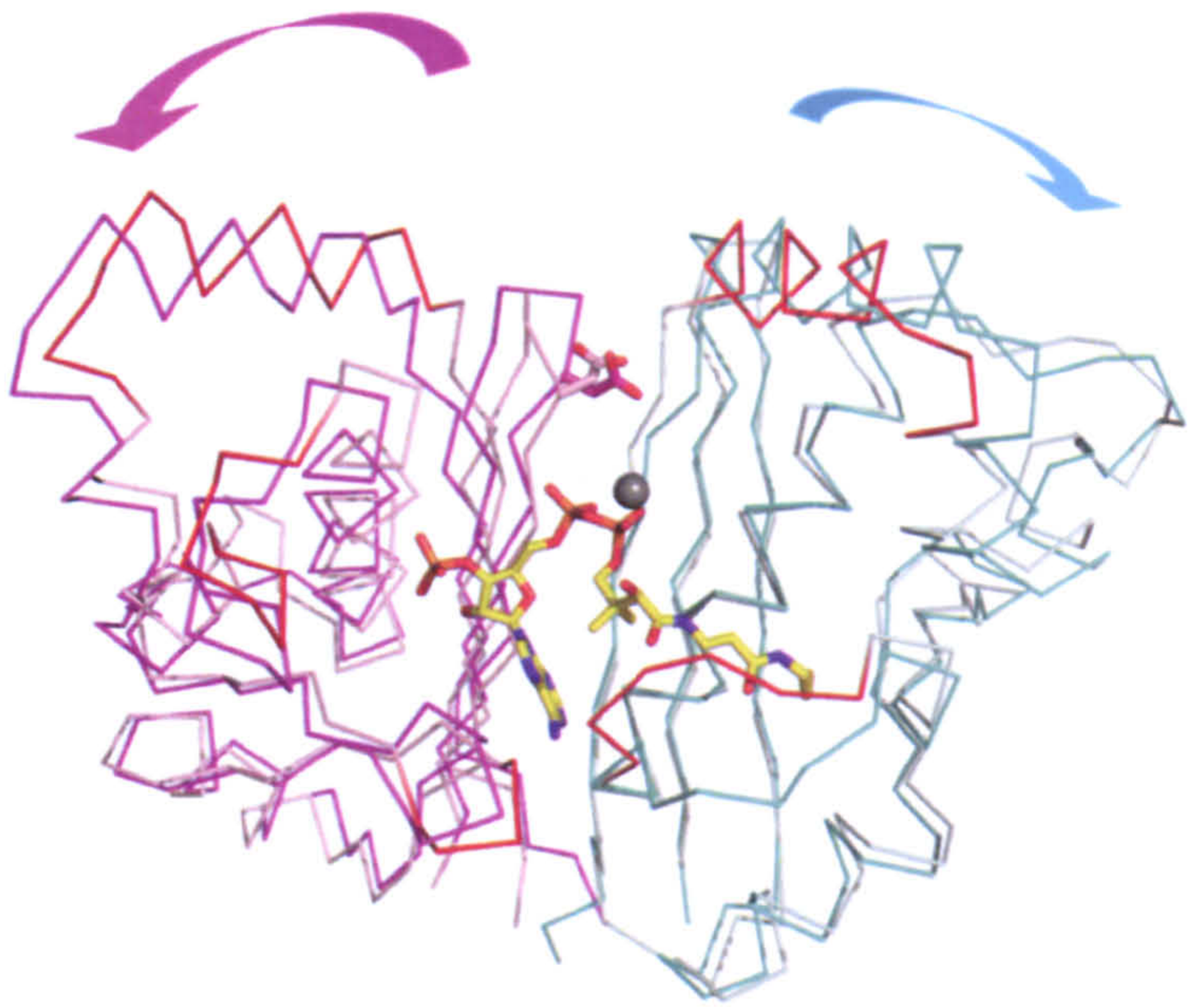


Figure 4.15 Superimposition between *S. coelicolor* holo- (in magenta and cyan) and apo-AcpS (in pale colors), picturing that the holo-protein seems to assume a more open conformation. Monomers A and B are shown as ribbon representation. Mobile loops are coloured in red. CoA is shown as sticks and coloured according to atom type. Magnesium is shown as a grey sphere.

4.7 Multiple sequence alignment of analogue type I phosphopantetheinyl transferases

The result of a multiple sequence alignment, carried out firstly with protein BLAST [147] and then with ClustalW [148], with analogue type I phosphopantetheinyl transferases is shown in Figure 4.16.

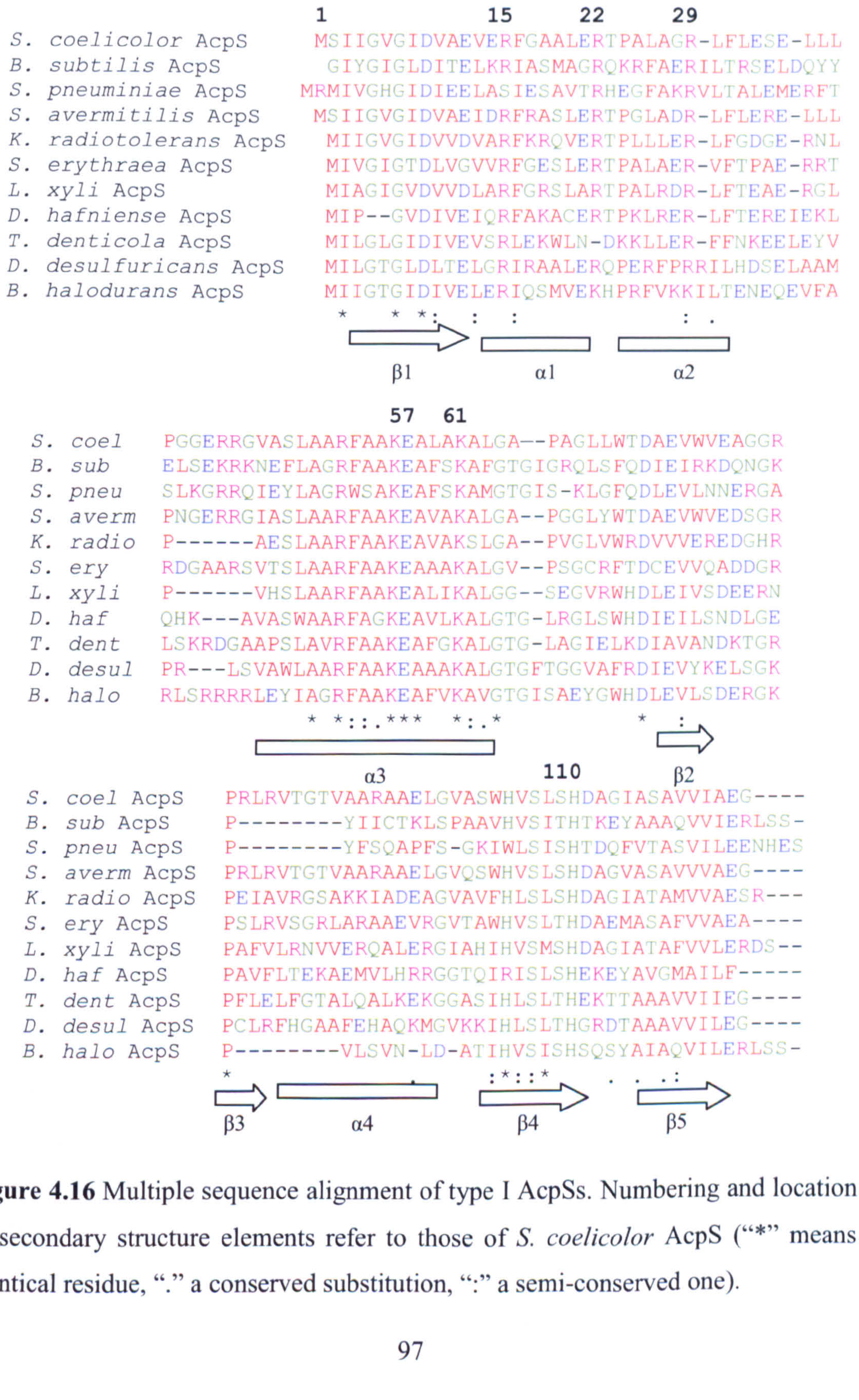


Figure 4.16 Multiple sequence alignment of type I AcpSs. Numbering and location of secondary structure elements refer to those of *S. coelicolor* AcpS ("*" means identical residue, "." a conserved substitution, ":" a semi-conserved one).

The various sequences are very similar to the *S. coelicolor* AcpS. 14 residues are completely conserved. The residues that were proposed to be important for ACP binding and for the catalysis [86] are Asp9, arginines 15, 22 and 29, Glu57, Lys61 and His105, and are very well conserved among the others type I AcpSs. While His110 is highly conserved, Asp111 is only partially conserved, and can be substituted by the similar glutamate residue or by a threonine, glycine or serine.

4.8 Comparison between *B. subtilis* and *S coelicolor* holo-AcpSs

The structure of the FAS *Bacillus subtilis* type I holo-AcpS [86] was superimposed on the *S. coelicolor* holo-AcpS, giving a RMSD of 1.35 Å between the Cα atoms of the 2 structures. Both structures are trimeric with similar folds, and there is a single subunit in the crystallographic asymmetric unit. The major difference between the folds is that there is an extra helix (helix α4) in the *S. coelicolor* structure (Figure 4.17). In the part of the *B. subtilis* sequence which corresponds to the helix α4 in *S. coelicolor* AcpS, a PCR error resulting in the mutation of a glutamine to proline in the crystallized protein was noted, which would preclude the formation of the extra helix in the *B. subtilis* AcpS [86].

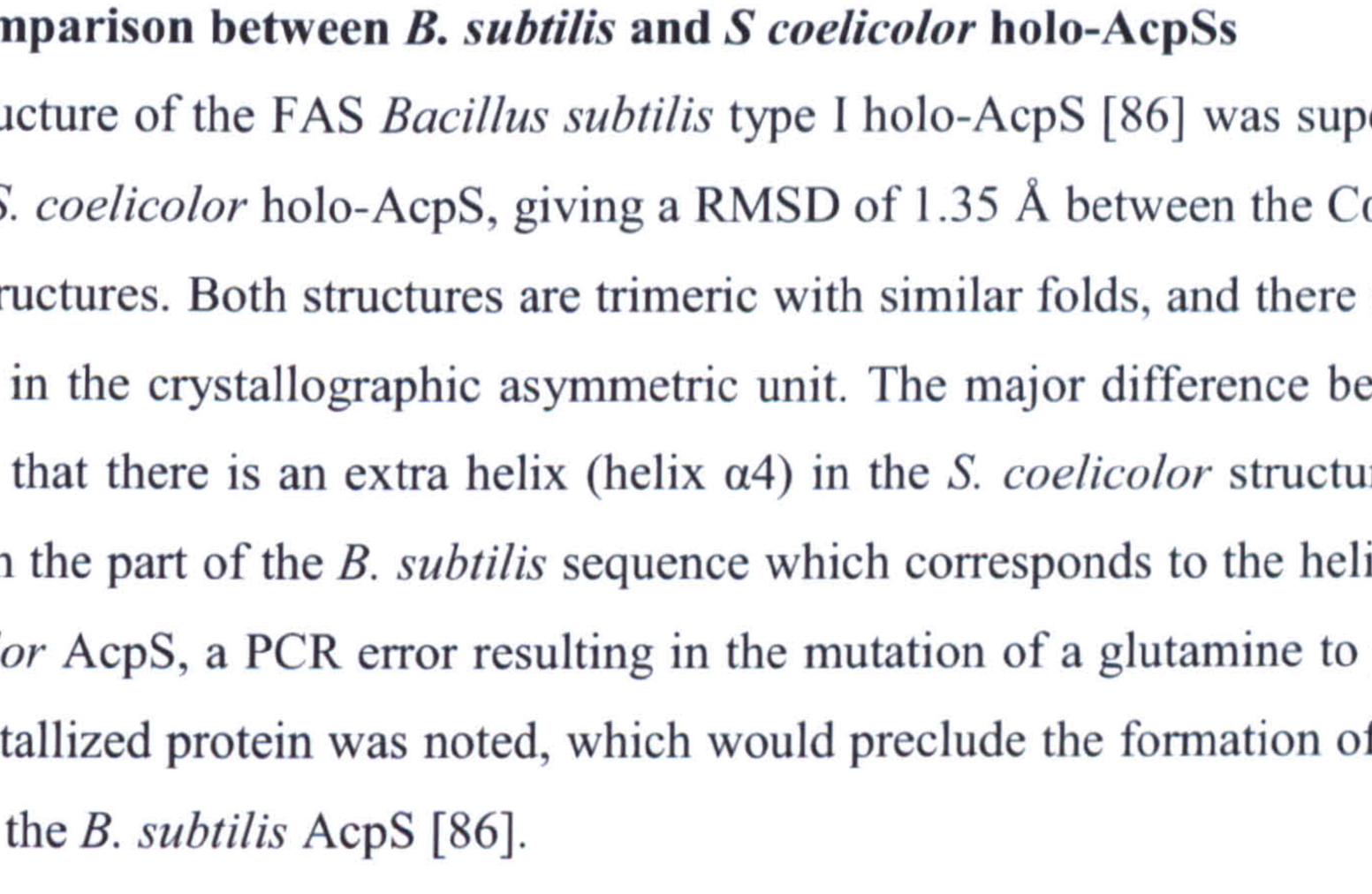


Figure 4.17 Cα superimposition between *S. coelicolor* (in magenta and cyan) and *B. subtilis* holo-AcpSs (in grey), shown as ribbons. Mobile loops are coloured in red. The proline introduced by PCR error in *B. subtilis* AcpS and CoAs are shown as sticks. *S. coelicolor* CoA is coloured in yellow, the *B. subtilis* one in grey. Magnesium is shown as a grey sphere, calcium as a green one.

The remaining differences are in regions surrounding the active site, interacting with the coenzyme CoA (Figure 4.18). The region that was identified as mobile and involved in binding the 3' phosphate, ML1 which is the loop at the end of helix 3, is replaced by a longer helix α3 in *B. subtilis* FAS AcpS, which joins on a longer central helix α3. ML3, which interacts with the adenine base, is in a similar conformation in both *S. coelicolor* and *B. subtilis* AcpS. ML2, which interacts with the phosphopantetheine chain, is in a different conformation. Comparing the location of the coenzymes in both structures, it is evident that they overlay almost perfectly from the tip of the adenine base up to and including the α-phosphate, and they diverge from the β-phosphate. However in the *B. subtilis* AcpS, a calcium ion is observed in the binding site of the magnesium required for catalytic activity due to the crystallization conditions [86]. Its coordination sphere includes Asp8, an oxygen from the alpha phosphate of CoA and 3 water molecules, and this is similar to the coordination observed for Mg^{2+} in the *S. coelicolor AcpS*. The final ligand for the calcium observed in the *B. subtilis* FAS AcpS is the side chain of Glu58, whereas the final ligand for the Mg^{2+} in *S. coelicolor* AcpS is an oxygen from the β-phosphate. In the *S. coelicolor* structure, the equivalent Glu57 side chain is in a different rotamer and creates a hydrogen bond with the phosphopantetheine chain.

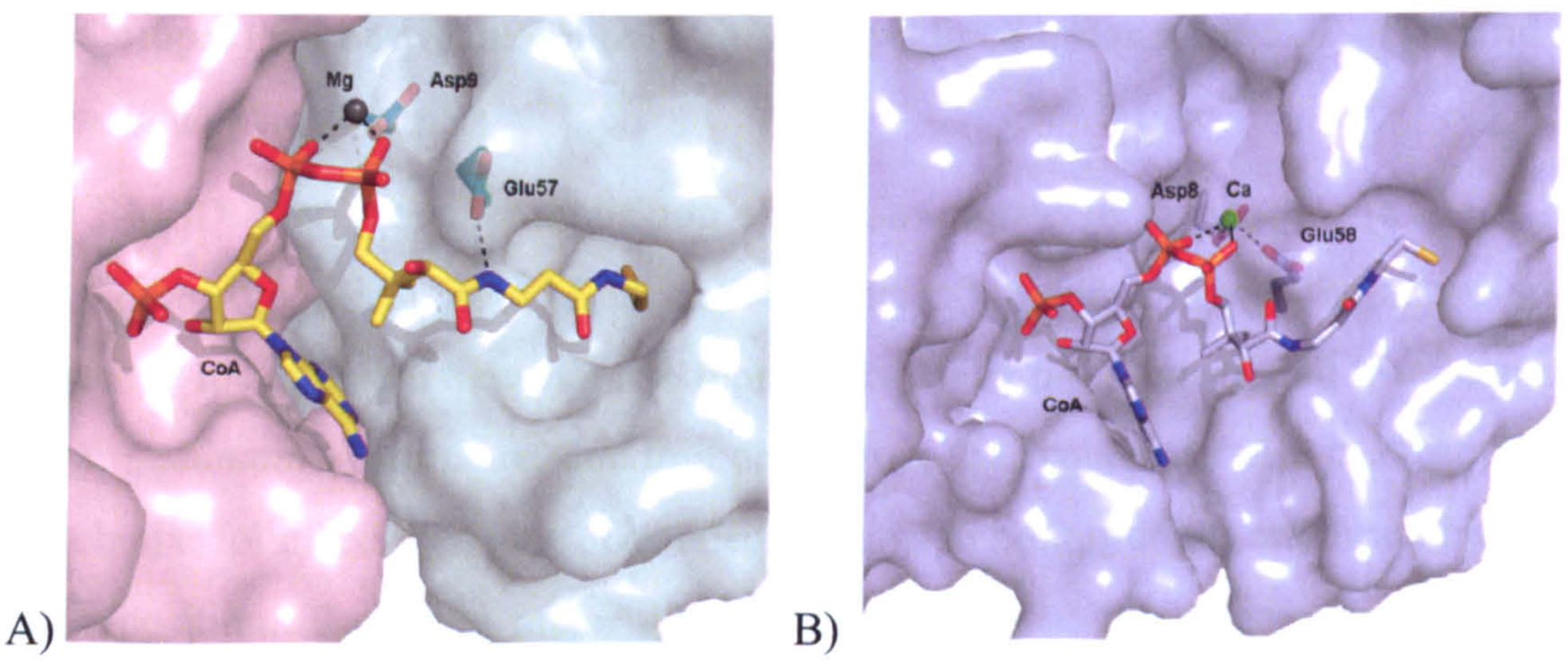


Figure 4.18 Surface representations for CoA binding pockets. A) *S. coelicolor* pocket shown in surface representation, with CoA shown as sticks and magnesium as a grey sphere. Monomers A and B are coloured in magenta and cyan. B) *B. subtilis* pocket shown in surface representation, with CoA shown as sticks and calcium as a green sphere.

It seems possible that the calcium observed in the *B. subtilis* structure, which is bigger than magnesium, affects the conformation of the alpha and beta phosphates and therefore the orientation of the subsequent phosphopantetheine chain in the *B. subtilis* AcpS structure. This is consistent with the observation that calcium allows catalysis in the *B. subtilis* enzyme, but at a much reduced rate. Whereas in the *S. coelicolor* structure the phosphopantetheine chain is buried in a hydrophobic pocket, and it is very well defined, in the *B. subtilis* structure it extends over the surface of the same monomer. The stretch defined earlier as ML2, which is mobile in the apo structure and forms the side of the pocket for the phosphopantetheine chain is in a different conformation in the *B. subtilis* holo structure, and the authors suggest it needs to move in order to accommodate the ACP. In the current holo structure, a similar conformational change occurs on CoA binding rather than ACP binding.

Finally, in the *B. subtilis* complex between AcpS and ACP [86], three AcpS arginines were proposed to be important for ACP binding: Arg14, Arg21 and Arg24. These three arginines are conserved in the *S. coelicolor* AcpS. As it will be discussed in Chapter 6 more in detail, Arg15, Arg22 and Arg29 can make salt bridges with negative residues located on the ACP.

4.9 Comparison between *S. pneumoniae* AcpS in complex with 3',5'-ADP and *S. coelicolor* holo-AcpS

The RMSD after Cα superimposition of the *Streptococcus pneumonia* [87] and *S. coelicolor* AcpSs is 1.72 Å. Both active proteins are trimers, with the active site located at the interface between two monomers. Helix α4 is substituted by a loop in the *S. pneumoniae* structure, due to a deletion in the sequence. ML1 is substituted by a helix, followed by a loop. The adenine rings in both proteins are located in the same position. The other major differences are located in the area identified as mobile in the *S. coelicolor* AcpS. In particular, while the region including ML1 is in the same conformation for the *B. subtilis* AcpS, the equivalent region in the *S. pneumoniae* structure is in a different conformation. ML3, which stacks against the adenine base of CoA, and its equivalent region in the *S. pneumoniae* protein, can be quite easily superimposed. The region equivalent to *S. coelicolor* ML2, which form one side of the pocket, is in a complete different conformation. In particular, *S.*

coelicolor ML2 seems to be very open, while the equivalent region in the *S. pneumoniae* structure, which is not determined in the presence of a phosphopantetheine arm, is very closed, in a conformation similar to that one assumed by the equivalent loop in the *B. subtilis* protein (Figure 4.19). This observation may be useful to explain the broader cofactor selection of *S. coelicolor* AcpS. This will be explored in more detail in chapter 5.

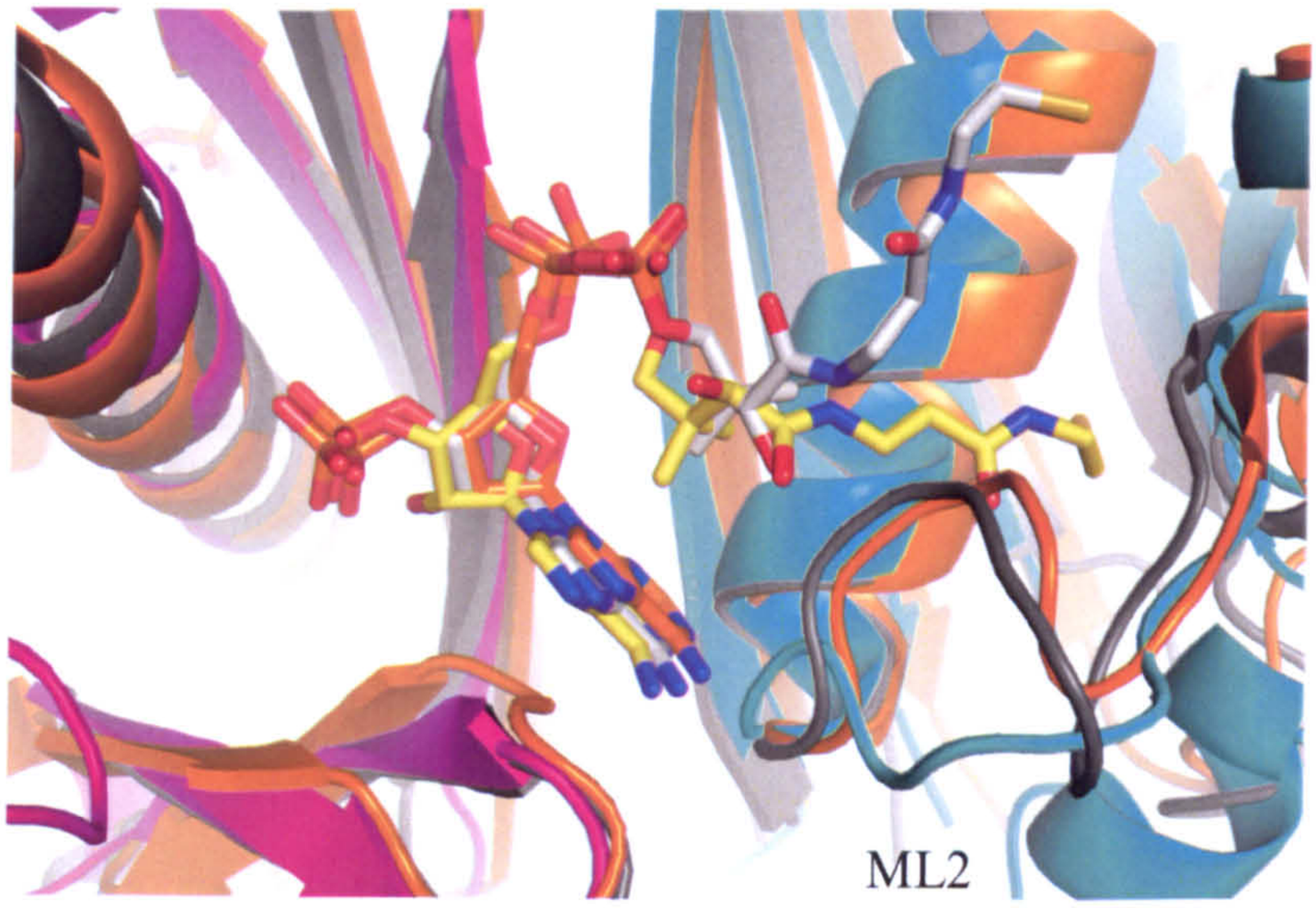


Figure 4.19 Comparison between ML2 of different proteins. *S. coelicolor* ML2 and CoA are shown in cyan and yellow respectively, *B. subtilis* ML2 equivalent region and CoA in grey, and *S. pneumoniae* ML2 equivalent loop and 3',5'-ADP in orange. The ribbon representation of *S. coelicolor* holo-AcpS monomers A and B is shown in magenta and cyan, of *B. subtilis* AcpS in grey and of *S. pneumonia* AcpS in orange.

With regard to the arginine residues that were proposed to be important for ACP binding [86], only *S. coelicolor* Arg22 and Arg29, are conserved in the *S. pneumoniae* AcpS, while Arg15 is not conserved. The interaction between *S. coelicolor* Arg15 and ACP Asp41, discussed in Chapter 6, will be proposed to be important for the catalysis. A different conformation of *S. pneumoniae* Arg46 could take the place of Arg15,

Finally, as it will be discussed in depth in Chapter 6 and 7, *S. coelicolor* His110 and Asp111 are proposed to deprotonate (in)directly the active serine on ACP. In the *S. pneumoniae* AcpS, His110 is conserved but is in a different conformation and a bit closer to the active site, but interacting with the 3' phosphate rather than the α-

phosphate of CoA. Analogue to *B. subtilis* AcpS, a threonine is found where the Asp111 is located.

4.10 Comparison between *S. coelicolor* and human holo-AcpS

Human AcpS [66] belongs to type II phosphopantetheinyl transferase family, a group that is represented by the Sfp protein from *B. subtilis* [78]. While *S. coelicolor* AcpS is 123 residues, the human one is longer - 315 amino acids. In addition, all type II enzymes show a trimeric arrangement, while the human AcpS is characterized by a 2 domain architecture, linked by a linker region. Each type II domain is similar to a single type I subunit from the trimer, thus the single proteins can be superimposed. Human AcpS has a longer N- and C-terminal in comparison to the *S. coelicolor* protein. The Cα superimposition between the holo-AcpS and the human one shows that the active site and the cofactor binding regions are very similar (Figure 4.20).

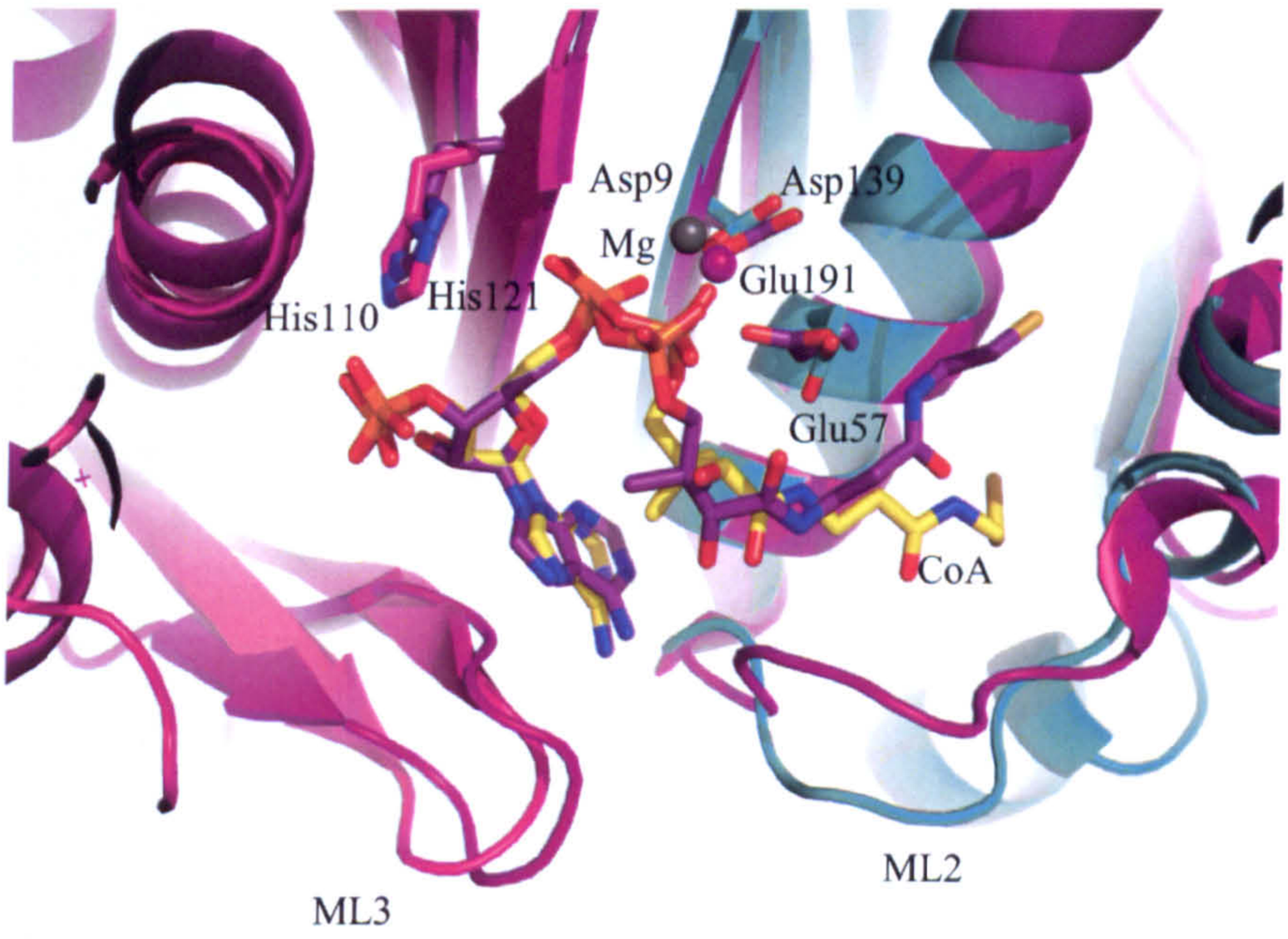


Figure 4.20 Superimposition between the human and *S. coelicolor* AcpS active sites. The ribbon representation of *S. coelicolor* AcpS monomer A and B is shown in magenta and cyan, the ribbon representation of human AcpS in purple. *S. coelicolor* CoA is coloured in yellow, the human one in purple. Magnesium ion is shown as a grey sphere for the *S. coelicolor* model, as a purple one for the human AcpS. Side chains for *S. coelicolor* AcpS are shown as sticks and coloured in magenta and cyan, for the human AcpS in purple.

CoA and magnesium are located at the interface between the 2 domains, with the ribose in the same 3'-endo conformation. ML1 is substituted by an α-helix, while ML3 is in a similar conformation. The adenine rings of CoAs are bound in a similar way, but the phosphopantetheine chains are bound in two different conformations. The two sulphur atoms are 6 Å apart. In addition, the cleft for the base seems to be wider in the human AcpS, and the human phosphopantetheine arm is located in a hydrophobic outer shell, provided by Ile140, Met142, Tyr184 and Trp187 [66], that is smaller than the *S. coelicolor* cavity. The region equivalent to ML2 is in a different conformation. *S. coelicolor* ML2 is more open, while the human one is in a intermediate conformation between the *B. subtilis* and the *S. coelicolor* loops (Figure 4.21). In the human model, the magnesium ion is coordinated to the two phosphates of CoA, a water molecule, an aspartate, equivalent to Asp9, and a glutamate residue, equivalent to Glu57.

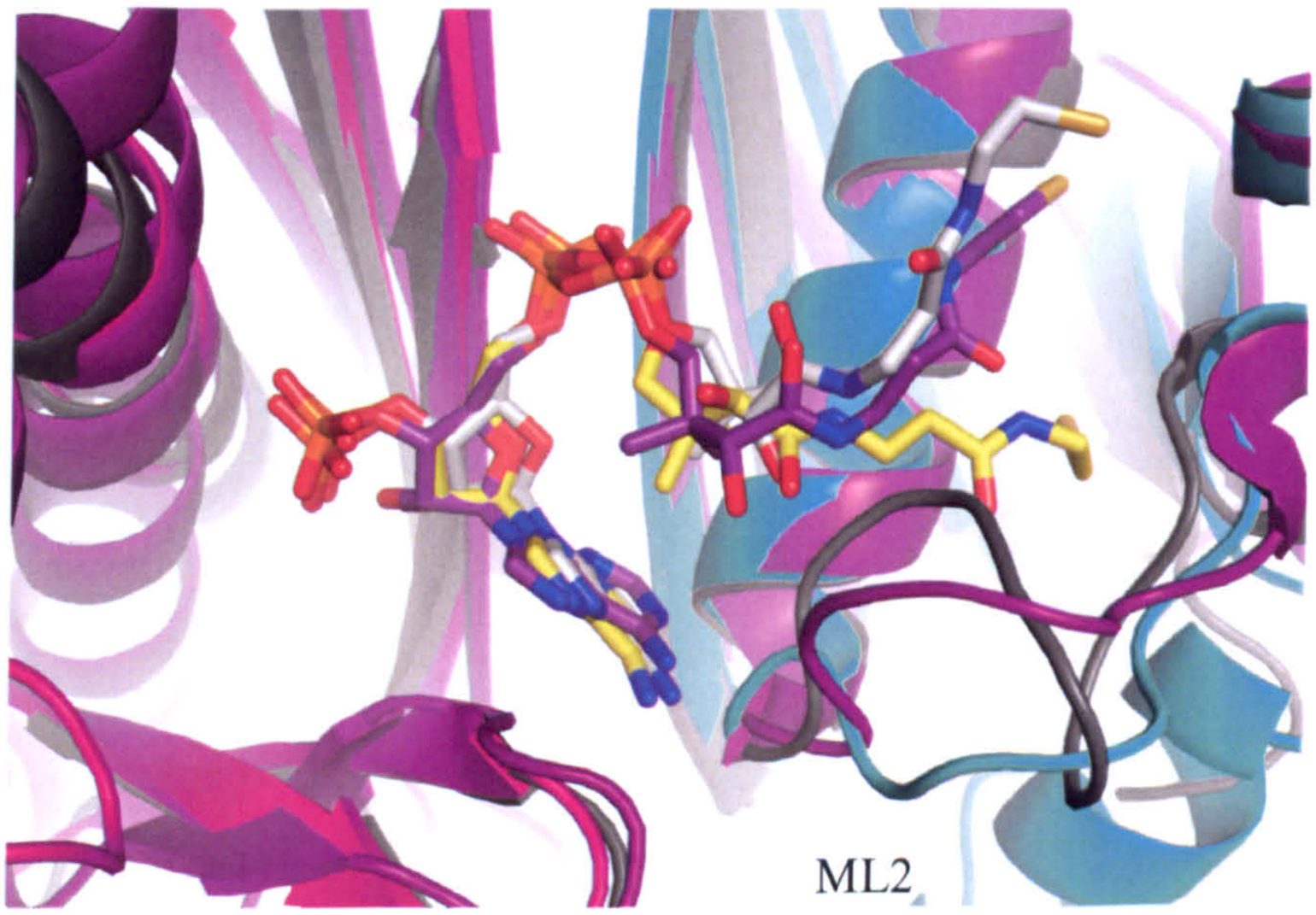


Figure 4.21 Comparison between different ML2 equivalent regions. *S. coelicolor* ML2 and CoA are shown in magenta and yellow respectively, *B. subtilis* ML2 equivalent region and CoA in grey, and the human ML2 equivalent loop and CoA in purple. The ribbon representation of *S. coelicolor* AcpS monomer A and B is shown in magenta an cyan, of *B. subtilis* AcpS in grey and of human AcpS in purple.

With regard to the interaction with the ACP (discussed in Chapter 6), even though *S. coelicolor* helix α1 is substituted by a loop in the human AcpS, Arg22 and Arg29 are conserved. Similarly to the *S. pneumoniae* structure, Arg15 is not conserved in the human AcpS, but human Arg65 can take its place. For the human AcpS it was proposed that a residue of the protein, Glu191, can directly deprotonate the serine in the ACP [66]. In the *S. coelicolor* model, the equivalent to Glu191 is Glu57, which interacts to the phosphopantetheine chain. In addition, *S. coelicolor* His110 is conserved, while Asp111 is replaced by a glutamine. This proposed mechanism is in great contrast to what proposed for type I phosphopantetheinyl transferase, where an activating water is believed to initiate the reaction deprotonating the ACP serine [86]. The presence of an activating residue for this type II enzyme shows strong parallels with the mechanism that will be proposed for the *S. coelicolor* AcpS (Chapter 6 and 7), and can be the starting point for finding an explanation to its promiscuity.

Chapter 5

Investigating the cofactor promiscuity of AcpS

In this chapter, the crystal structure of the *S. coelicolor* AcpS in complex with the acetyl-CoA is presented. A comparison with the holo-AcpS is shown, and hypotheses about the cofactor promiscuity in the active site are discussed. ITC was also used to determine the affinity constants for CoA and acetyl-CoA binding.

5.1 AcpS in complex with acetyl-, propionyl- and malonyl-CoA

S. coelicolor AcpS is considered to be a promiscuous enzyme, capable of working with modified versions of CoA [74]. Indeed, it is used by colleagues in the Department of Chemistry for routinely preparing a range of holo- and acyl- forms of ACP. As can be seen from Figure 5.1, CoA is surrounded by a very well defined hydrophobic cavity, which allows even more room for extended version of the cofactor.

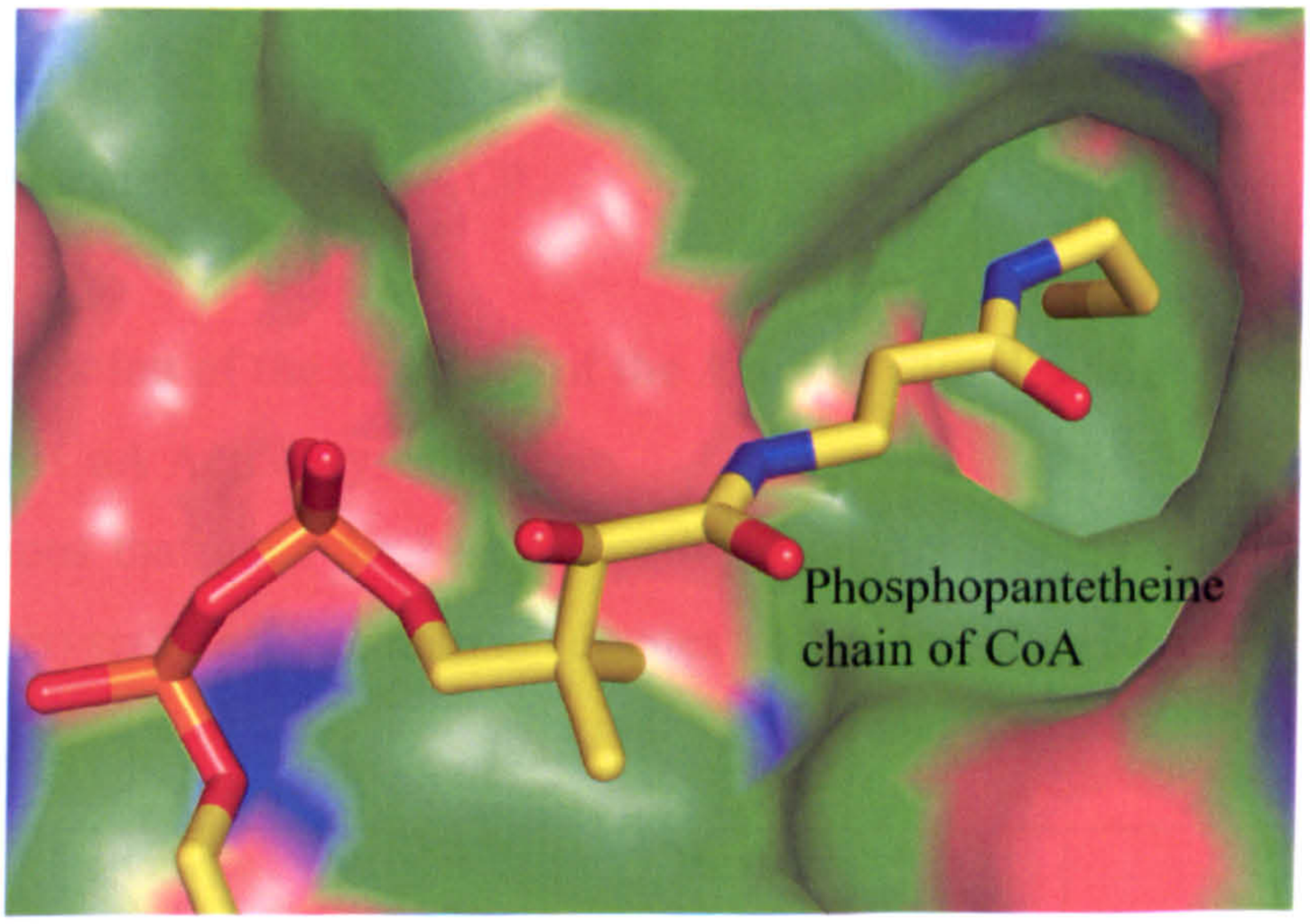


Figure 5.1 AcpS hydrophobic cavity. Charged surface representation of the hydrophobic cavity that surrounds the phosphopantetheine chain. The neutral regions are coloured in green, while the positive and negative areas are shown in blue and red respectively. CoA is shown as sticks and coloured according to atom type.

Whilst it has been shown that this *S. coelicolor* AcpS can transfer an aceto-acetyl-, hexanoyl- and benzoyl-CoAs [74], from the crystal structure of the holo-AcpS, there seems not to be enough space for very long or aromatic activating groups. To experimentally explore the promiscuity in terms of substrate transfer, we tried to solve the structure of the AcpS in complex with modified CoA of a reasonable length and shape, such as acetyl-, malonyl- and propionyl-CoAs (Figure 5.2).

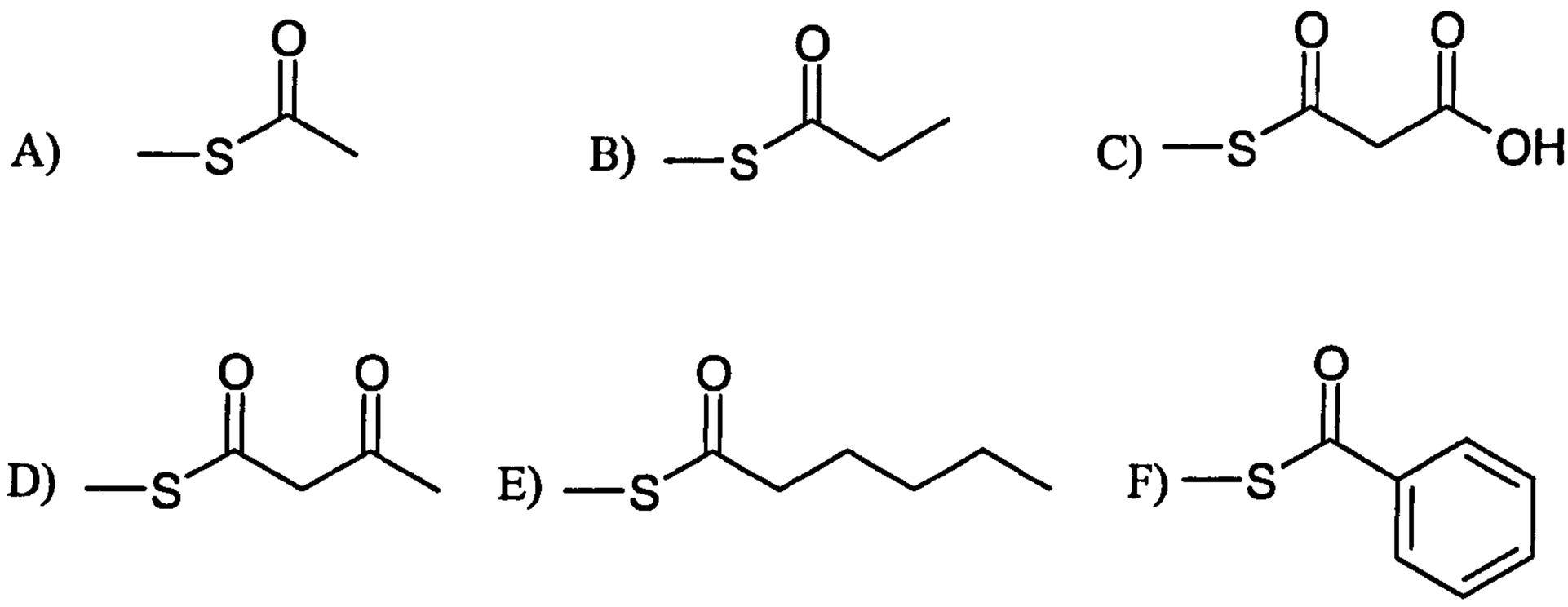


Figure 5.2 Cofactor promiscuity of *S. coelicolor* AcpS. Structure of the acetyl (A), propionyl (B), malonyl (C), acetoacetyl (D), hexanoyl (E) and benzoyl (F) groups that are covalently attached to the phosphopantetheine arm of CoA and were proved to be transferred to the ACP.

5.2 Determination of the structure of the AcpS in complex with acetyl-CoA

5.2.1 Acetyl-CoA AcpS crystallization

The conditions that gave the best diffracting holo-AcpS crystals (table 5.1) were screened to crystallize the protein in complex with acetyl-CoA (chapter 3.8.1).

Salt	Buffer	PEG
0.2 M lithium sulphate	0.1 M NaCacodylate pH 6.5	25% PEG 2K MME
0.1 M potassium thiocyanate	0.1 M NaCacodylate pH 6.5	15% PEG 4K

Table 5.1 Crystallization conditions for AcpS in complex with acetyl-CoA.

Optimization trays were prepared for both conditions. Salt concentrations of 0.1 M, 0.2 M and 0.3 M were tested, and the PEG concentration was varied in 2% (for PEG 2K, between 19-27% w/v) or 1% steps (for PEG 4K, between 12-18% w/v). Regular and triangular crystals, measuring about 0.7×0.6×0.3 mm, were obtained at 18°C after a couple of days (Figure 5.3), using a protein concentration of 10 mg/mL and in the presence of 5 mM of acetyl-CoA.

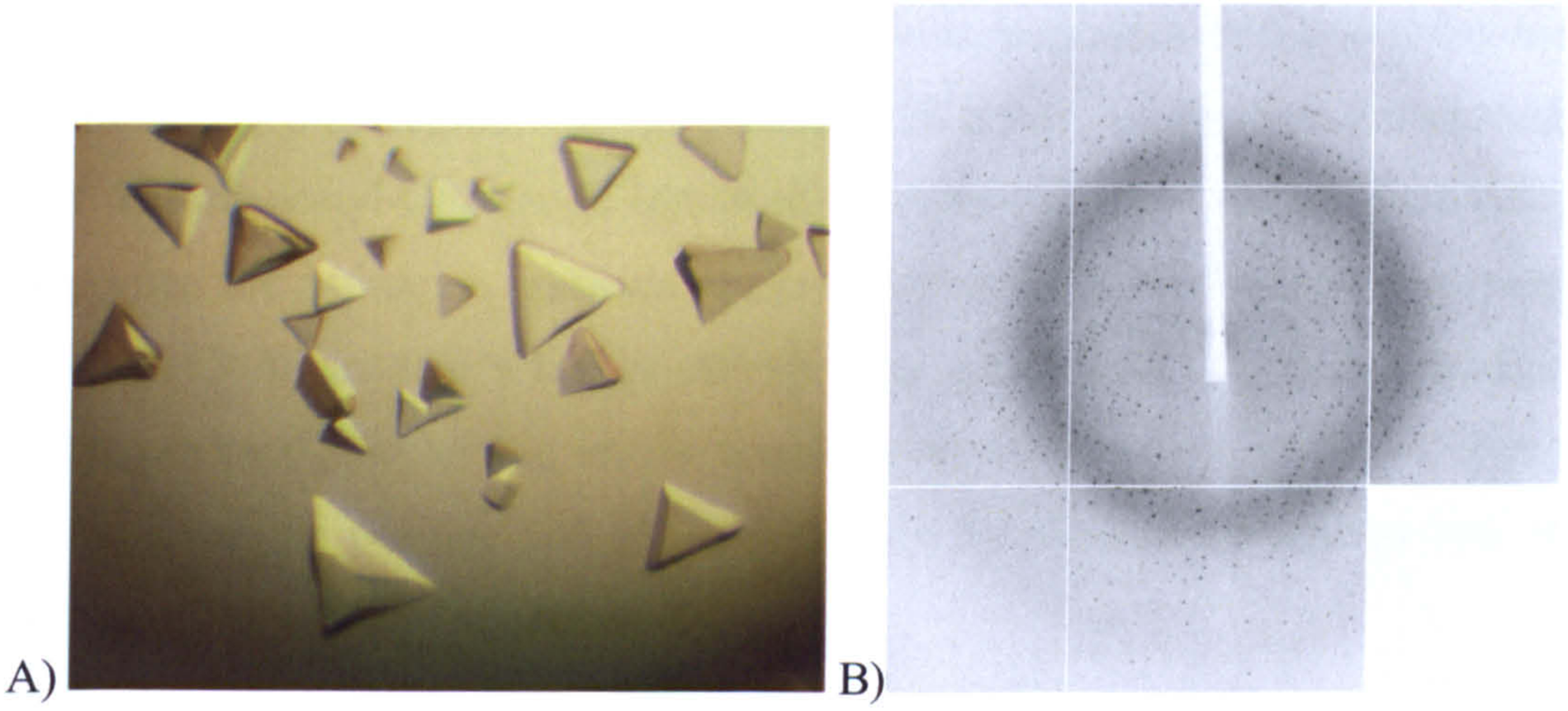


Figure 5.3 Acetyl-CoA AcpS crystals. A) Typical acetyl-CoA AcpS crystals obtained in 0.2 M lithium sulphate, 25% PEG 2K MME, 0.1 M sodium cacodylate pH 6.5. B) An example of a diffraction image. One of the CCDs of the detector was not working during data collection, and this region was excluded during integration.

5.2.2 Data collection and processing

Several crystals were tested at beamline 10.1 at the Synchrotron Radiation Source (SRS), Daresbury, and IO2 and IO3 at the Diamond Light Source (DLS). The best diffracting crystal was obtained in 0.2 M lithium sulphate, 25% PEG 2K MME, 0.1 M sodium cacodylate pH 6.5, with the presence of acetyl-CoA at the concentration of 5 mM. The reservoir plus 15% v/v glycerol was used as cryoprotectant. Data were collected at the IO2 station. This crystal diffracted to an approximately resolution of 1.5 Å and belonged to the cubic $P2_13$ space group with a=b=c=72.8 Å. The data processing was carried out using the HKL2000 [135] and CCP4 suites of programmes [143].

5.2.3 Molecular replacement and refinement

Similarly to the holo-AcpS, one molecule per asymmetric unit was found. The structure was solved by molecular replacement using Phaser [136, 137] and the holo-AcpS as search model. In addition, in order to verify that the right cofactor was present, Phaser was run initially without the presence of acetyl-CoA, in order to have an unbiased map. The model contains 123 residues, 1 magnesium ion, 1 molecule of acetyl-CoA with 0.75 occupancy, 1 of CoA with 0.25 occupancy, 1 glycerol, 1 sulphate ion and 237 water molecules. Inspection of the electron density in the acetyl-CoA region, initially refined with a full occupancy cofactor, and the fact that the acetyl-CoA stock contained a contamination of the standard cofactor, suggested the additional presence of CoA (Figure 5.4). As it will be demonstrated by ITC, CoA has also a higher affinity than the acetyl-CoA. In this model, the difference map was used to set the correct occupancy for both cofactors, by trying to minimize the positive and negative density. Finally, the occupancy was set to 0.25 for CoA and 0.75 for acetyl-CoA.

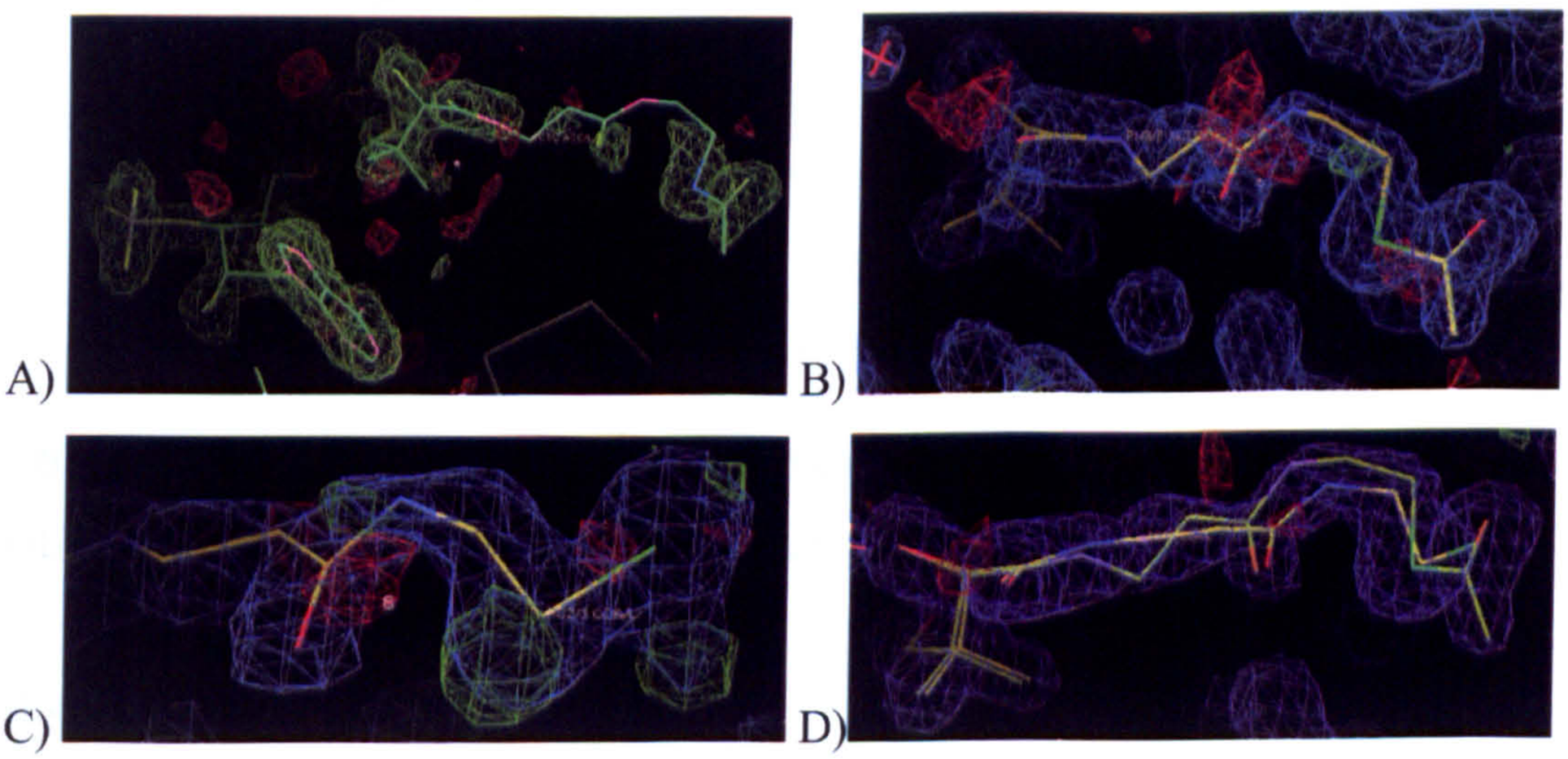


Figure 5.4 Suggested presence of CoA in the acetyl-CoA AcpS active site. A) Unbiased difference map (DELFWT, DELPHWT) after refining the model without cofactor (acetyl-CoA from the final model is shown for comparison). The electron density maps (FWT, PHWT) and difference maps (DELFWT, DELPHWT) after refinement with acetyl-CoA (B) or CoA (C) are then shown. D) Maps related to the final model, with acetyl-CoA and CoA in the active site. The electron density map is contoured at 1.2σ, the difference map at 3σ. Cofactors are shown as bonds.

In addition, the electron density for the small loop containing Asp111 was not very clear: an apo-like and a holo-like form of this loop could be identified. The refinement led to a structure model with R-factor an R_{free} values of 0.20 and 0.23 respectively. Statistics for data processing and refinement are shown in Table 5.2.

	Acetyl-CoA AcpS
Wavelength (Å)	0.953
Resolution (Å)	50-1.56 (1.62-1.56)
Space group	P 2_13 a=b=c=72.8 Å
N unique reflection	18659
Completeness (%)	100 (100)
Redundancy	9.4 (9.1)
I/σI	32 (2.5)
R merge	0.057 (0.381)
Mosaicity (°)	0.18
R_{free}	0.23
R-factor	0.20
RMS bond (Å)	0.012
RMS angle (°)	1.533

Table 5.2 Statistic data for AcpS in complex with acetyl-CoA. Data related to the highest resolution shell are reported between brackets.

5.2.4 Geometry and B-factor analysis

The geometry and the conformation were checked using MolProbity [145, 146]. The Ramachandran plot is presented in Figure 5.5. 97.8 % of the residues are in the most favoured region, while Ala65 and Asp111 are in the allowed region. No residues were recognized as having unusual rotamers. An analysis of the average B-factor of the main chains was also performed with Baverage [143]. The average B-factor is 17 $Å^2$, while the average temperature factor for the mobile areas, except for ML2, is higher: 22 $Å^2$ (MH1), 30 $Å^2$ (ML1) 16 $Å^2$ (ML2) and 24 $Å^2$ (ML3). However, the average B-factor for ML2 is higher than the neighbouring regions.

The loop containing the Asp111 has also a higher average B-factor: 25 $Å^2$. The temperature factor analysis and the comparison with the wild type AcpS are shown in Figure 5.6. The average B-factor of the AcpS in complex with acetyl-CoA is lower than the wild type one, but its pattern is very similar, with the mobile loops characterized by a higher temperature factor.

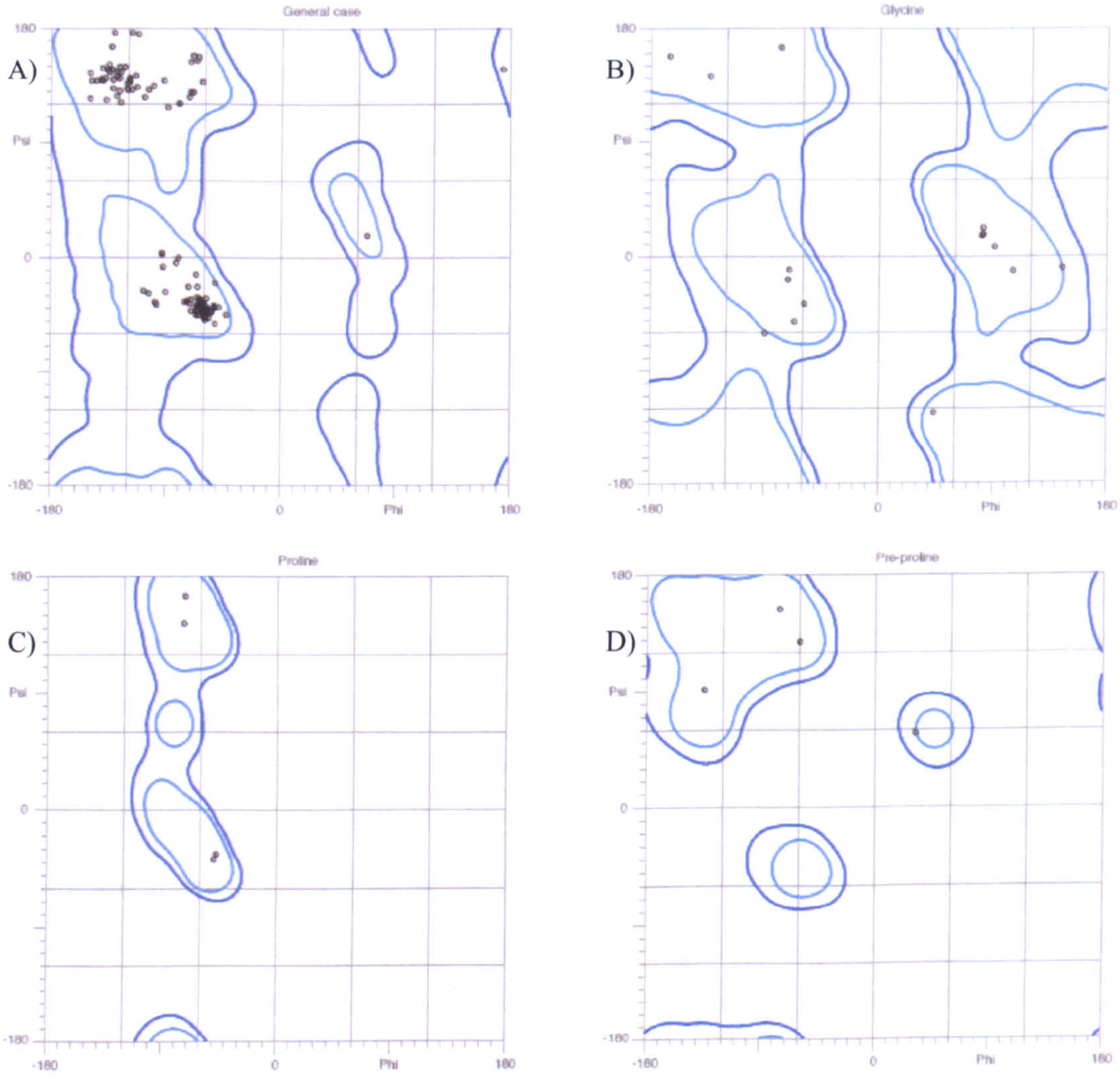


Figure 5.5 Ramachandran plot related to the acetyl-CoA AcpS. All residues are in the favoured regions with the exception of Ala65 and Asp111, that are located in the allowed region. A) general case; B) glycine; C) proline; D) pre-proline.

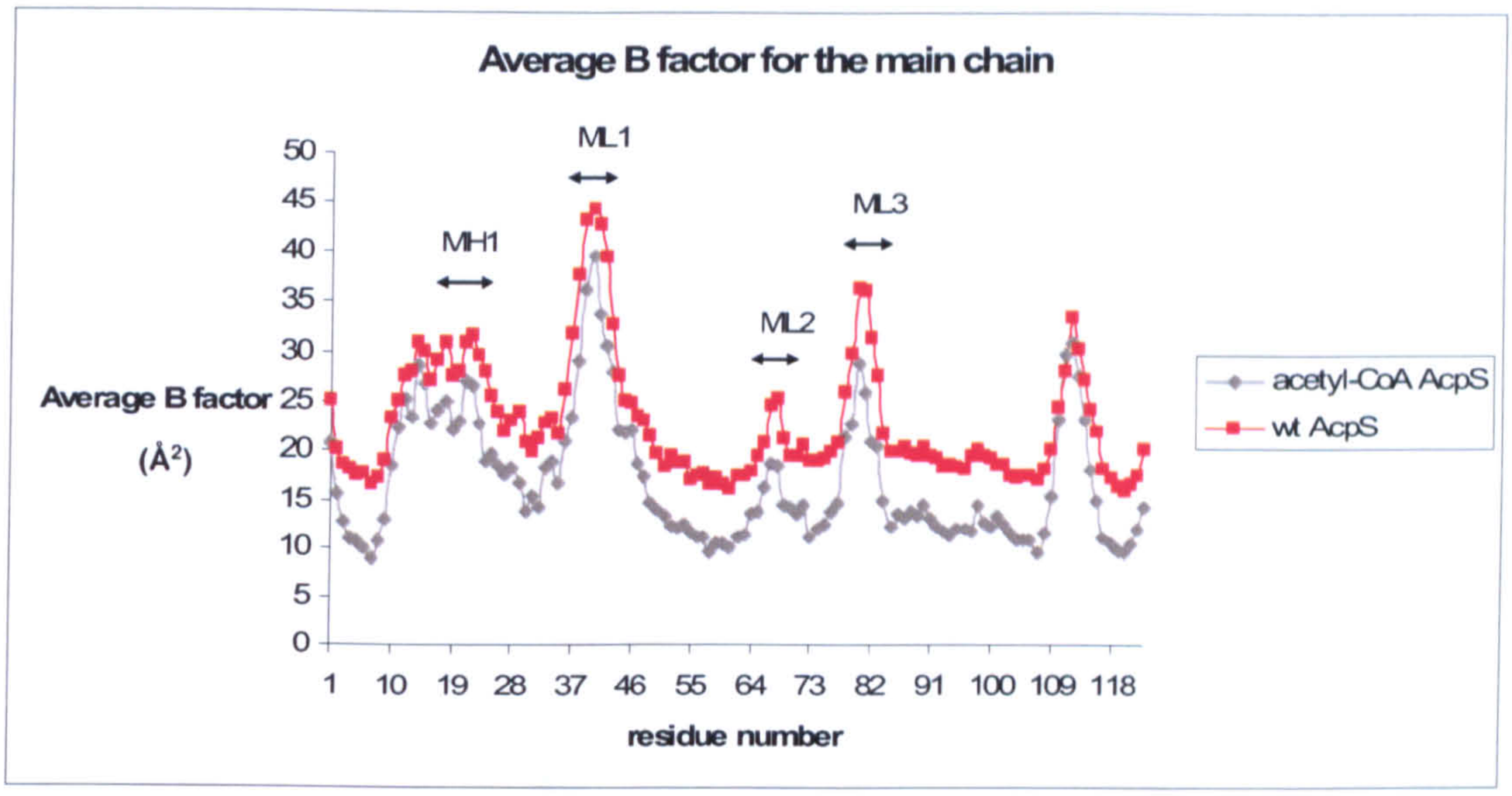


Figure 5.6 Acetyl-CoA AcpS B-factor analysis. Average B-factor values for the main chain of AcpS in complex with acetyl-CoA. The B-factor values reflect the mobility of the mobile loops, and the pattern is similar to that of the holo-AcpS.

5.3 Analysis of the crystal structure of AcpS in complex with acetyl-CoA

The model for the acetyl-CoA AcpS is very similar to the holo-AcpS. The RMSD after Cα superimposition between the AcpS in complex with CoA and the one with acetyl-CoA is 0.20 Å. In addition to the presence of CoA, the electron density for the small loop containing the Asp111 is not clear: both an apo-like and a holo-like forms of this loop can be seen, but no explanation for this disorder was found. The density for the acetyl-CoA is very clear, especially for the extra acetyl group. All mobile loops have the same conformations in comparison to the protein in complex with CoA. The magnesium ion is 1.9 Å away from the cation of the holo structure, and similarly to the loop containing the Asp111, no particular interaction or conformational changes were noticed to cause this displacement. The magnesium coordination sphere is different in comparison to the holo-AcpS, but no explanation for the displacement was found. Whilst in the holo structure it was coordinated to both phosphates of CoA (with a distance of 2 Å and 1.9 Å respectively), to the carboxyl group of Asp9 (2.1 Å) and 3 water molecules (within 2.2 Å of distance), the magnesium in the acetyl-CoA structure is 3.3 Å and 3.2 Å far from the α- and β-phosphates, 2.6 Å from the main chain oxygen of Val10, 3.2 Å from the carboxyl group of Asp111, and 3.4 Å from Asp9. In addition, the ion is coordinated to waters Z230 (the equivalent to Z12, 2.9 Å far away) and Z232 (equivalent to Z183,

3.1 Å distant), but there does not seem to be any equivalent to water Z16 (Figure 5.7). Water Z232 is quite far from its equivalent water Z183. Other two waters, Z152 or Z231, within 4.2 Å, could complete its coordination sphere.

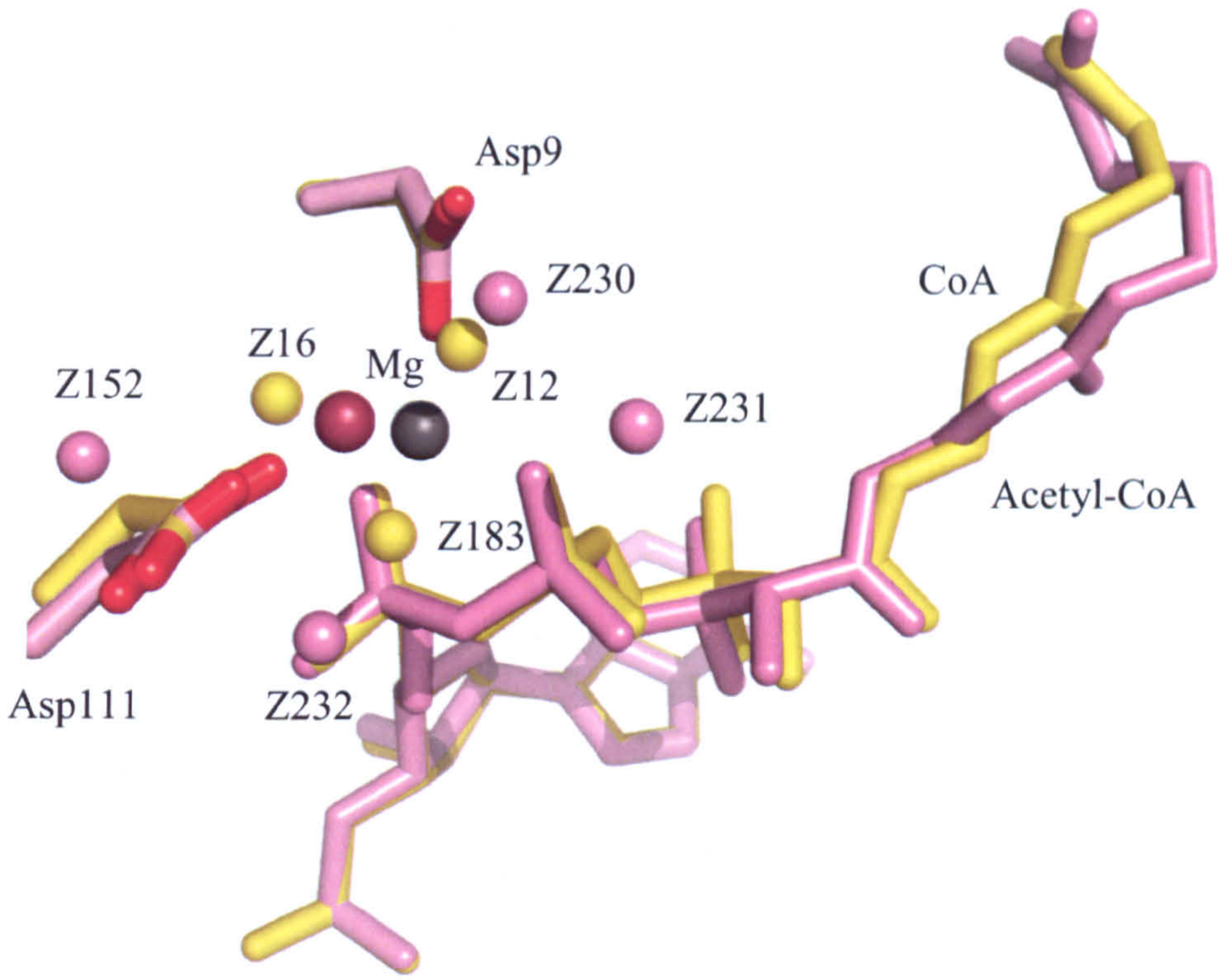


Figure 5.7 Comparison between the holo-AcpS and acetyl-CoA AcpS active sites. Amino acidic residues are shown as sticks, water molecules and magnesium ions as spheres. CoA, residues and waters belonging to the holo-AcpS are coloured in yellow, while acetyl-CoA, waters and residues belonging to the acetyl-CoA structure are shown in pink. The magnesium in the holo-AcpS is coloured in grey, the ion in the acetyl-CoA in pink.

The α-phosphate is within hydrogen bonding distance with the carbonyl group of Asp9, the side chain of Lys61, the hydroxyl group of the Ser109, a water molecule and the main chain oxygen of His110, while the β-phosphate with Asp9 and 3 water molecules. The oxygen of PO9 is held in place through hydrogen bonds provided by waters Z209 and Z107, and all the other interactions present in the CoA bound protein are conserved. Whilst the His110 retains its interactions, Asp111, which only interacts with magnesium via a water in the holo structure, makes a direct interaction with the magnesium ion in the acetyl-CoA AcpS. The two cofactors are

disposed in the same position until the PC5 atom, and then the other atoms are displaced up to 2.4 Å. The acetyl-CoA seems to adopt a more contracted conformation between PC3 and PS1 atoms rather than to cause evident conformational changes on ML2 (Figure 5.8). The sulphur atom of CoA and the carboxyl group of the acetyl-CoA are just 0.65 Å apart. All the residues that form the hydrophobic cavity that contains the phosphopantetheine arm are disposed in the same way and with the same conformations. Superimposing both trimers using a Cα rigid superimposition of monomers A, ML2 in the acetyl-CoA protein results to be displaced of about 0.65 Å in comparison to the ML2 of the protein in complex with CoA. It was supposed that ML2 can move to accommodate extended CoAs. The addition of just 3 atoms (a carbonyl and methyl group) does not seem to cause any dramatic conformational changes.

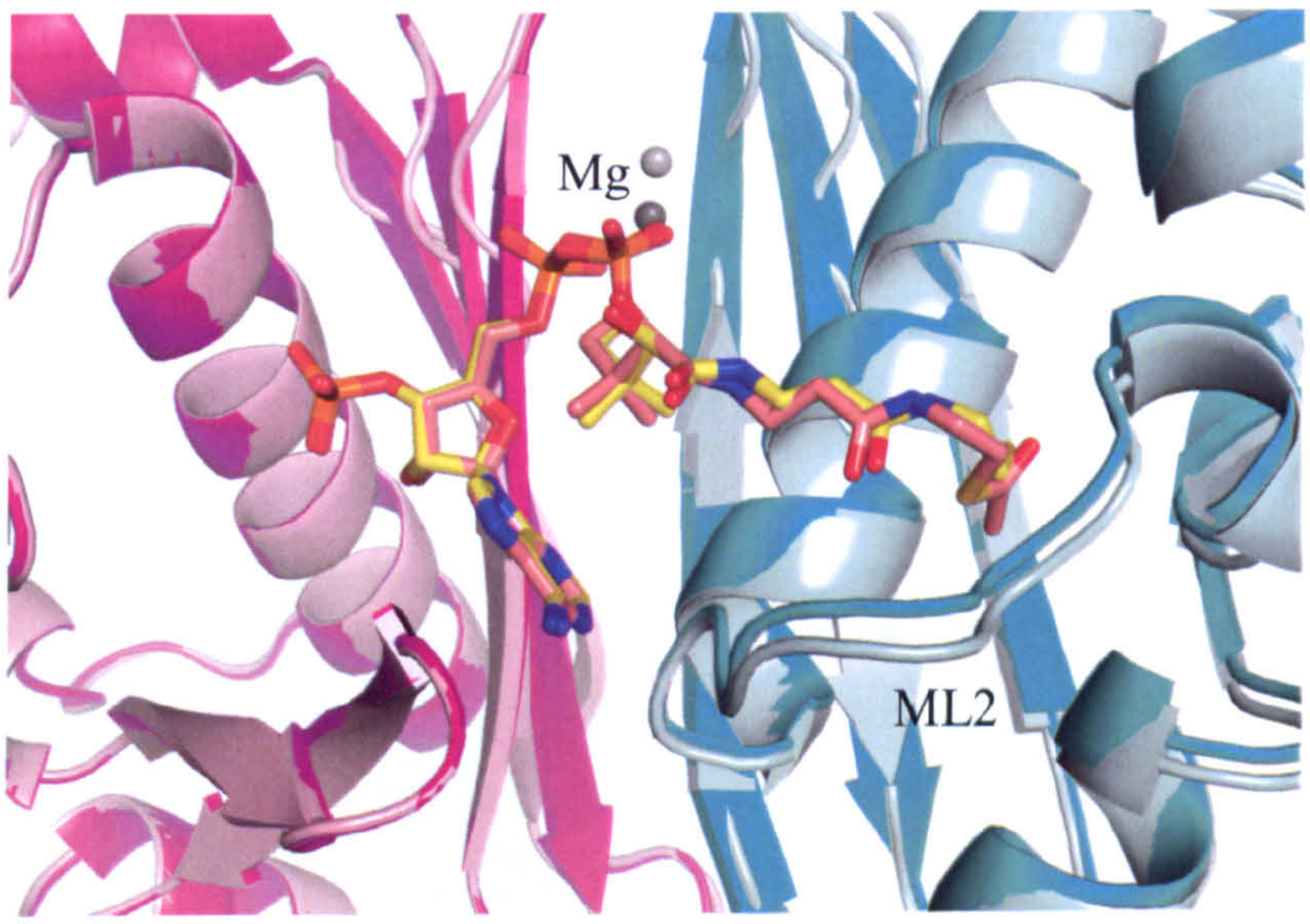


Figure 5.8 Conformation of ML2 after CoA and acetyl-CoA binding. Cartoon representation of holo-AcpS, shown in magenta and cyan, and AcpS in complex with acetyl-CoA, pictured in pale colours. CoA (in yellow) and acetyl-CoA (pink) are shown as sticks.

5.4 Crystallization of AcpS in complex with malonyl- and propionyl-CoA

Crystallization trials, using the same crystallization conditions for the acetyl-CoA AcpS (Table 5.1), were set after incubating AcpS with solutions of frozen or fresh propionyl- (5 or 10 mM final concentration) and malonyl-CoA (5 mM). Crystals appeared after 1-2 days at 18 °C (Figure 5.9). These crystals had very similar dimension and triangular shape of the other holo-AcpS proteins. Several datasets were collected for each complex at beamline 10.1 at the Synchrotron Radiation Source (SRS), Daresbury, and IO2 and IO3 at the Diamond Light source (DLS). All crystals belonged to the cubic $P2_13$ space group, with similar unit cell dimensions to the other holo-AcpS proteins. Table 5.3 shows the data collection statistics for several datasets. To check the presence of the appropriate cofactor, the structures were solved by molecular replacement using the holo-AcpS as search model with or without the presence of the cofactor CoA. Unfortunately, all the structures solved resulted in holo forms of the protein. Inspection of the difference map did not reveal any evidence of a longer density corresponding to propionyl- or malonyl-CoA. Consequently, the cofactors were analyzed by mass spectroscopy in order to investigate their stability and grade of purity in solution.

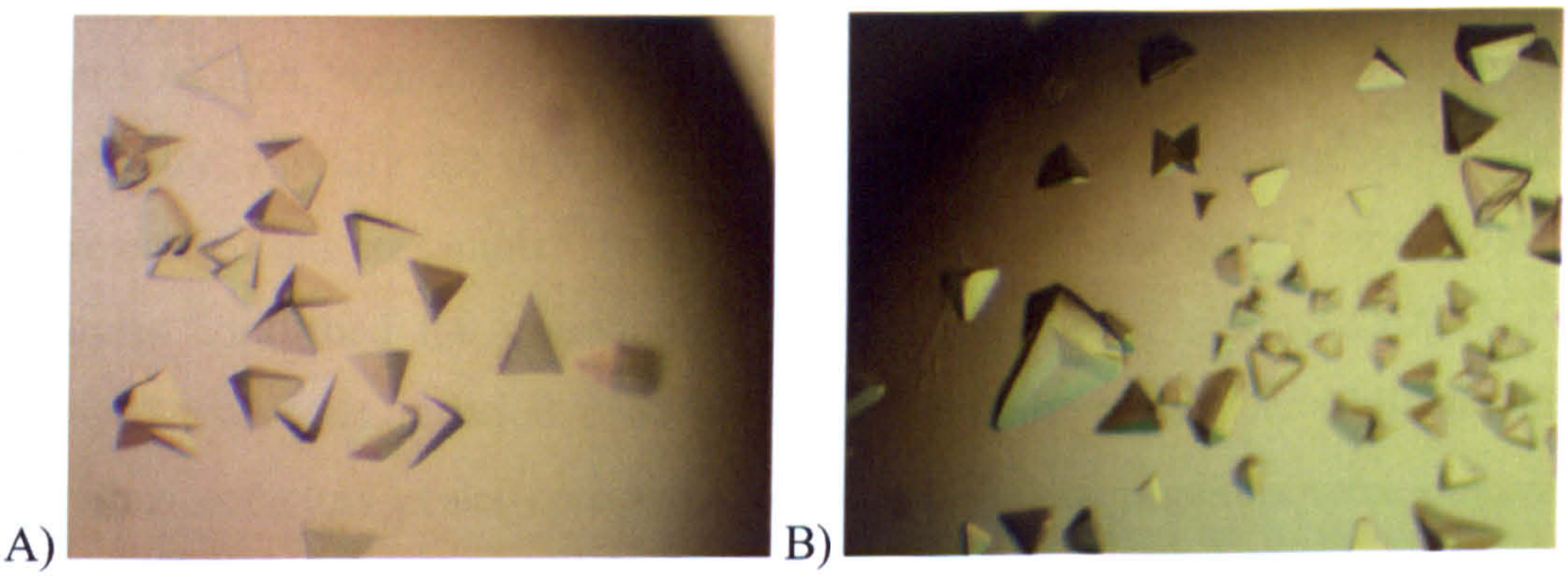


Figure 5.9 Crystals of AcpS in the presence of malonyl- (A) and propionyl-CoA (B). A) Crystals obtained in 0.2 M lithium sulphate, 0.1 M NaCacodylate pH 6.5, 25% PEG 2K MME. B) Crystals obtained in 0.1 M potassium thiocyanate, 0.1 M NaCacodylate pH 6.5, 15% PEG 4K.

Cofactor	propionyl-CoA (10 mM, frozen)	malonyl-CoA (5 mM, frozen)	malonyl-CoA (5 mM, fresh)
Wavelength (Å)	0.92	0.953	0.92
Resolution (Å)	50-1.8 (1.86-1.8)	50-1.7 (1.76-1.7)	50-2.3 (2.38-2.3)
Space group	P 2_13 a=b=c=72.7 Å	P 2_13 a=b=c=72.9 Å	P 2_13 a=b=c=72.7 Å
N unique reflection	12206	14515	5702
Completeness (%)	100 (100)	100 (100)	100 (100)
Redundancy	10.8 (10.6)	9.7 (9.4)	8.6 (9)
I/σI	23.1 (2.5)	20.6 (2.2)	31.3 (2.1)
R merge	0.11 (0.78)	0.10 (0.74)	0.07 (0.7)
Mosaicity (°)	0.65	0.85	0.79

Table 5.3 Data collection statistics for AcpS incubated with propionyl- or malonyl-CoA. Data corresponding to the highest resolution shell are in brackets.

5.5 Mass spectroscopy analysis of cofactors

All cofactors incubated with the AcpS were analyzed by mass spectroscopy (kindly carried out by Chris Arthur, Department of Chemistry) to check their state in solution (chapter 3.7.5). CoA, acetyl-CoA and malonyl-CoA were bought from Sigma-Aldrich, while propionyl-CoA was a kind gift of the Department of Chemistry. Initially the crystallization trials and the mass spectroscopy analysis were carried out on samples that were stored at -20°C and were 1 month old. This analysis confirmed the presence of CoA, acetyl- and propionyl-CoA, but also showed that a degradation process was present. In particular, while in all the previous cofactors peaks corresponding to CoA contaminations were present, no peak related to malonyl-CoA was found. The malonyl-CoA stock seemed to be all degradated to CoA and acetyl-CoA. So, new crystallization trials and mass analysis were carried out with fresh new cofactors immediately after resuspension in water and prior the incubation with the protein. All spectra are quite salty and different deprotonation states for each cofactor are present (Figure 5.10).

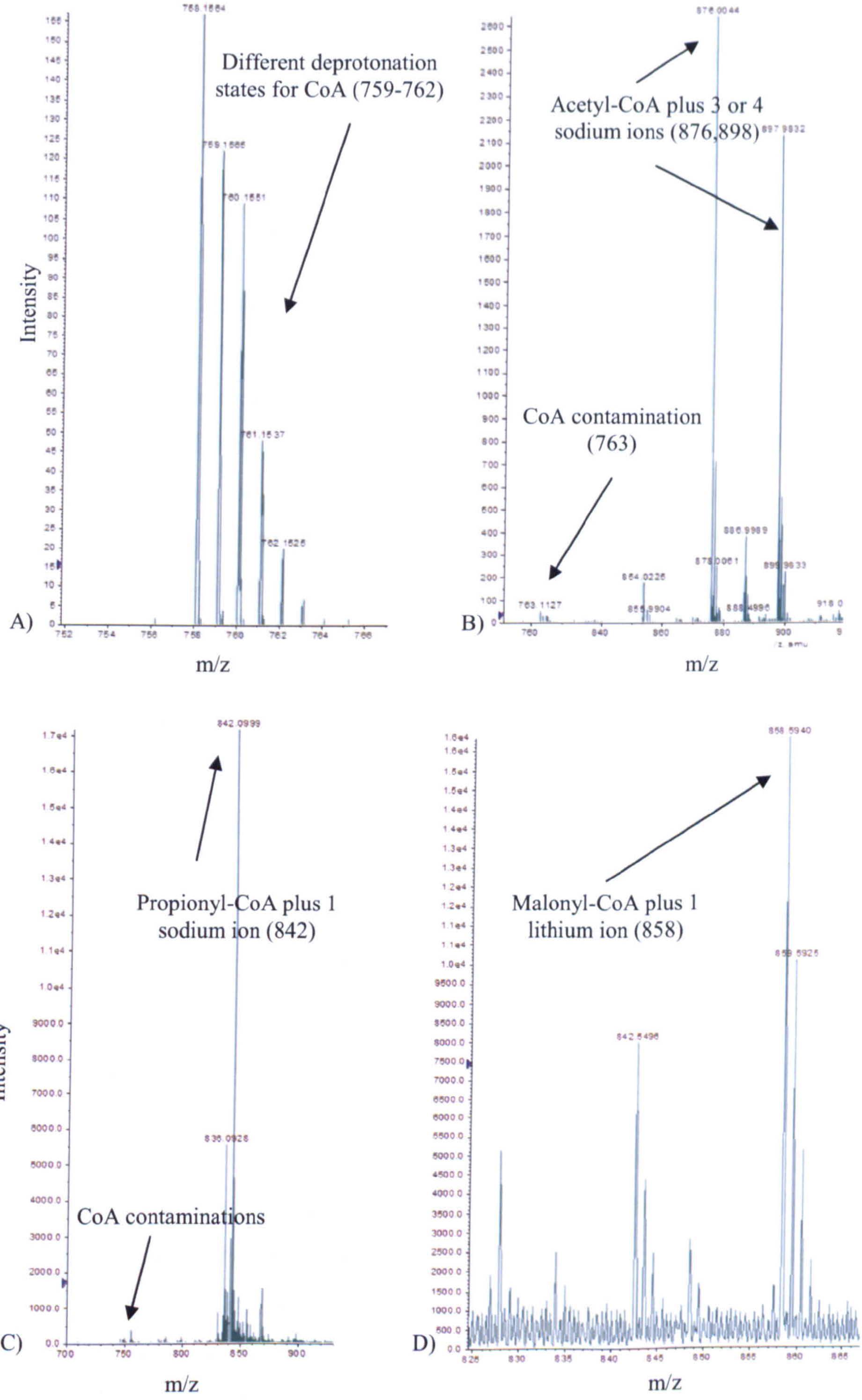


Figure 5.10 Mass spectra of CoA (A), acetyl- (B), propionyl- (C) and malonyl-CoA (D). All spectra are quite salty. The presence of all cofactors is confirmed. All errors are within ± 0.9 kDa.

CoA is in the form of a sodium salt hydrate, with a purity of 96% and a calculated mass of 767.53 Da. From its mass spectrum, it is possible to notice that several peaks are present, from 758.15 to 762.15 Da, reflecting different deprotonation states of CoA. Acetyl-CoA is sold as a sodium salt, 95% pure, with a calculated mass of 809.57 Da. It is enzymatically prepared by reacting CoA with acetyl phosphate and phosphotransacetylase, and the product is then purified by ion exchange chromatography. The mass spectrum shows the presence of CoA as a contamination, with other peaks corresponding to acetyl-CoA, and 3 or 4 sodium ions are present. For malonyl-CoA, it is in the form of a lithium salt, with a purity of 90% and a calculated mass of 853.58 Da. The mass spectrum confirms the presence of this cofactor with a lithium ion bound, but also shows that the degradation process is still present even if the stock was fresh. Finally, the spectrum related to propionyl-CoA, characterized by a weight of 823.6 Da, shows its presence with a sodium ion bound.

5.5.1 Instability of cofactors

As proved by mass spectroscopy analysis, a considerable degree of degradation affects especially the propionyl- and malonyl-CoA in solution, and contaminations of CoA are present in the other stocks. Datasets collected from crystals obtained after incubation with frozen malonyl-CoA did not result in any structure with the modified cofactor bound. This result did not change when fresh cofactor was used. Several hypotheses can explain the fact that just CoA is bound to the protein. There can be a sort of self hydrolization of the cofactor catalyzed by the AcpS, but this seems unlikely since acetyl-CoA is observed in the crystal structure and it was proved that these modified CoAs can be transferred to the ACP [74]. These cofactors could not be as stable as indicated by the manufacturer. In particular, malonyl-CoA seems to be stable just for a few days. The crystals obtained after incubation with fresh malonyl-CoA were frozen within a week after their appearance. To avoid degradation, it would be very useful to freeze them as soon as they appear. In addition, CoA, that is present as a contamination, could have an affinity that is much tighter than malonyl- and propionyl-CoAs, similarly to what happens in the acetyl-CoA structure. Even though incubation with 10 mM of propionyl-CoA was not successful, it would be interesting to set new trays with

higher malonyl- or propionyl-CoA concentration. Moreover, purification of these compounds by HPLC could also be a good way to get rid of the contaminants. Finally, it would be useful to try to soak the longer cofactors into apo-AcpS crystals, and investigate whether the conformation of the apo-protein will change after cofactor binding.

5.6 Determination of K_d for CoA and acetyl-CoA by ITC

ITC was used to determine the binding constant for AcpS and CoA or acetyl-CoA (chapter 3.9). In the acetyl-CoA structure, CoA seems also to be present as contamination, and the occupancy for the two cofactors were set to 0.75 (acetyl-CoA) and 0.25 (CoA), suggesting a weaker binding for the acetyl-CoA, that is in higher excess in the stock used. ITC brilliantly confirmed what was predicted by the structures. Preliminary experiments indicated that an initial CoA concentration of 120 μM should be used to obtain an accurate K_d, avoiding saturation that occurs at a higher cofactor concentration. K_d and the stoichiometry for the two cofactors are reported in Table 5.4.

	CoA	Acetyl-CoA
K_d (nM)	380 ± 20	2400 ± 100
n	1.2 ± 0.1	0.94 ± 0.07
ΔH (cal/mol)	-4000 ± 400	-1600 ± 200

Table 5.4 Binding affinity for the holo-AcpS and the acetyl-CoA AcpS. K_d, stoichiometry and difference in enthalpy are shown for CoA and acetyl-CoA binding to the wild type AcpS.

As expected, it was found that a molecule of CoA/acetyl-CoA binds to a monomer of AcpS (3:3 ratio). The value for n is affected by about a 10% error that could be due to pipetting error or to experimental errors in the determination of the protein or cofactor concentration. K_d for CoA is about 6.4 times tighter than for the acetyl-CoA. The weaker binding for the longer cofactor is also evident from the shape of the ITC graphs. Whilst the curve related to CoA has almost a rectangular shape and reflects a high value of the ITC constant c, the one relating to the acetyl-CoA starts to resemble to a horizontal line with a lower c value (Figure 5.11). Finally, the

change in entropy for CoA binding is greater than for the acetyl-CoA. A blank without the presence of the protein was also run to determine the background heat of dilution of the cofactor. A straight line, indicating no binding as expected, can be seen in Figure 5.12.

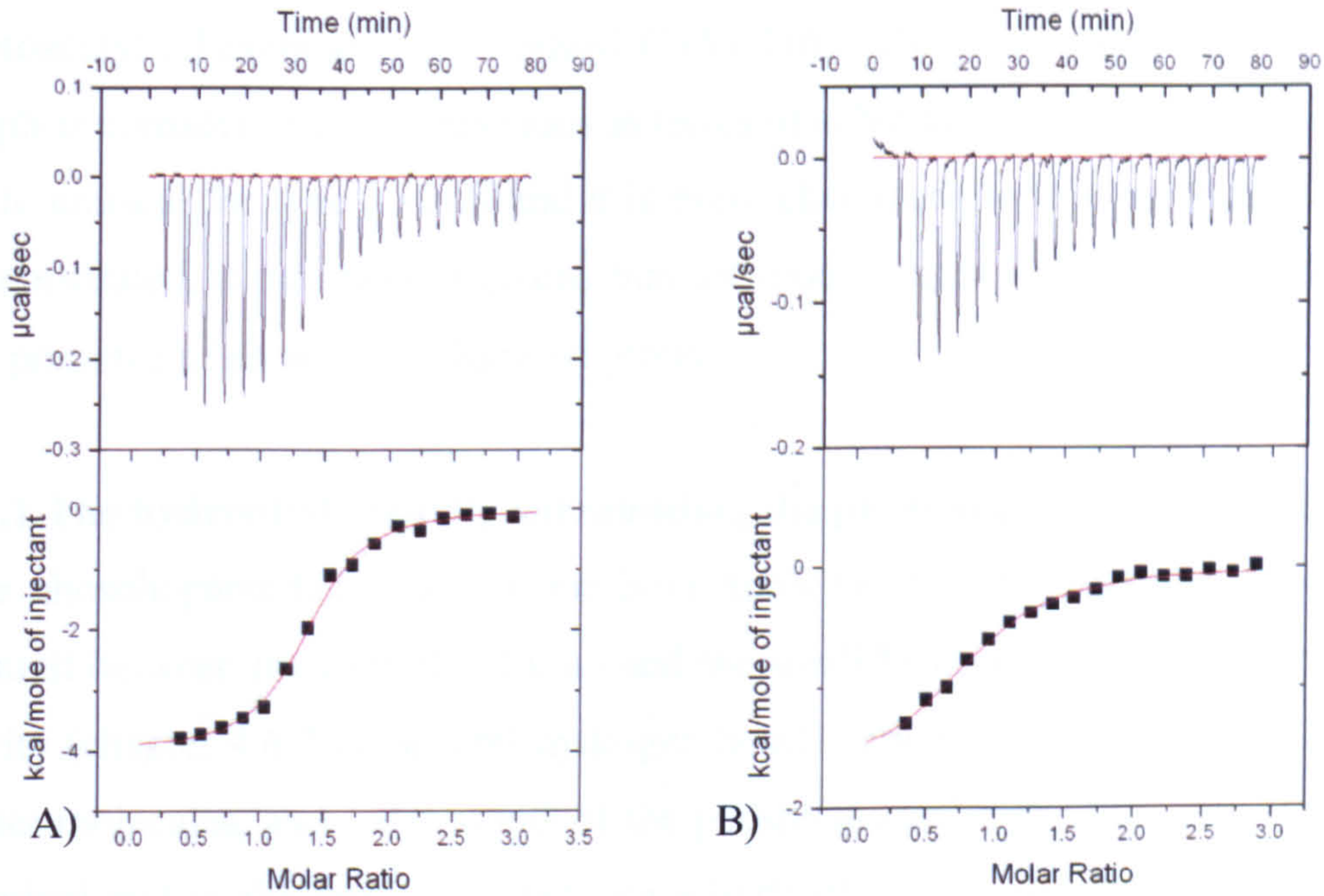


Figure 5.11 ITC data showing CoA and acetyl-CoA binding to wild type AcpS. The affinity for the acetyl-CoA (B) is lower than that for CoA (A).

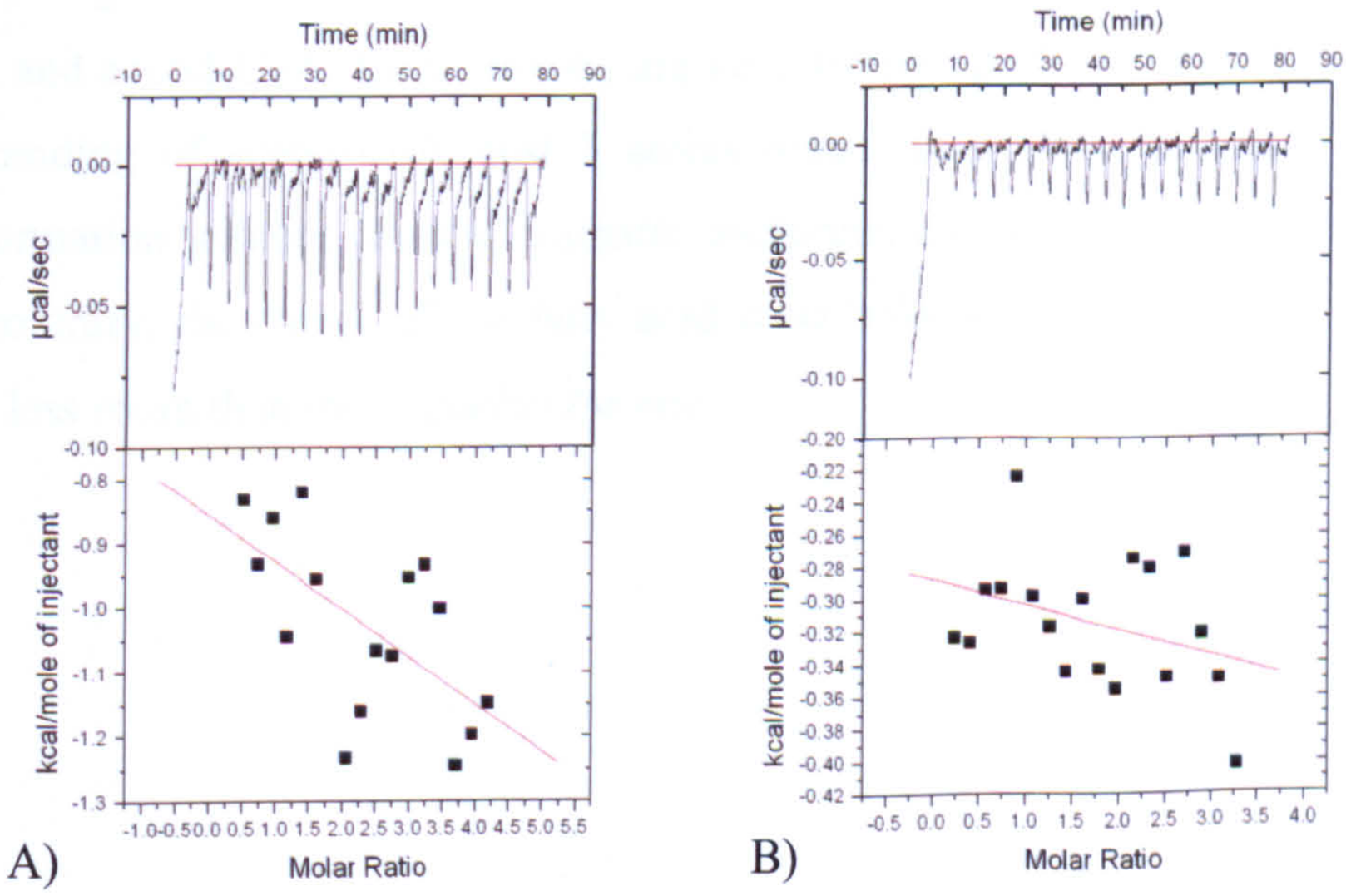


Figure 5.12 Blanks for CoA (A) and acetyl-CoA binding (B).

5.7 Cofactor promiscuity of *S. coelicolor* AcpS

Russell *et al.* [16] proved that the *S. coelicolor* AcpS can transfer several acyl-phosphopantetheine chains from acyl-CoA to the apo-ACP. In addition to the natural CoA, this protein can promote the phosphopantetheinylation of ACP by accepting acetyl-, malonyl-, methylmalonyl-, propionyl-, n-butyryl-, isobutyryl-, acetoacetyl-, hexanoyl- and benzoyl-CoAs [16]. Therefore, type I *S. coelicolor* AcpS is consider to be promiscuous in terms of substrate binding, a property that is quite unusual for type I AcpS and it is more characteristic of type II enzymes. As demonstrated, it can transfer chains that are much longer than CoA, and also with the presence of an aromatic benzoyl group.

5.7.1 The hydrophobic cavity surrounding the phosphopantetheine chain

The phosphopantetheine arm in the holo-AcpS lies in a hydrophobic cavity on a subunit between the central helix α3 and the small helix α5. At the entrance of the cavity (chapter 4.6.2.1), several hydrogen bonds between protein residues or some water molecules, and polar atoms of the phosphopantetheine chain, are made. The terminal end of the arm is located into a hydrophobic cavity that is sealed by the apolar side chains of residues that form ML2. The cavity that surrounds the phosphopantetheine chain of CoA in the holo-AcpS is very well defined, and there seems to be more room for extra atoms. Water molecules are excluded from the cavity. Figure 5.13 shows the cavities surrounding the phosphopantetheine arms of CoA and acetyl-CoA. Both cavities are very similar, probably due to the fact that the binding of acetyl-CoA, just 3 atoms bigger than CoA, adopts a contracted conformation without causing dramatic conformational changes to the protein. On the contrary, the cavity of the fatty acid *B. subtilis* AcpS (Figure 4.18 B) allows even less room than the *S. coelicolor* one.

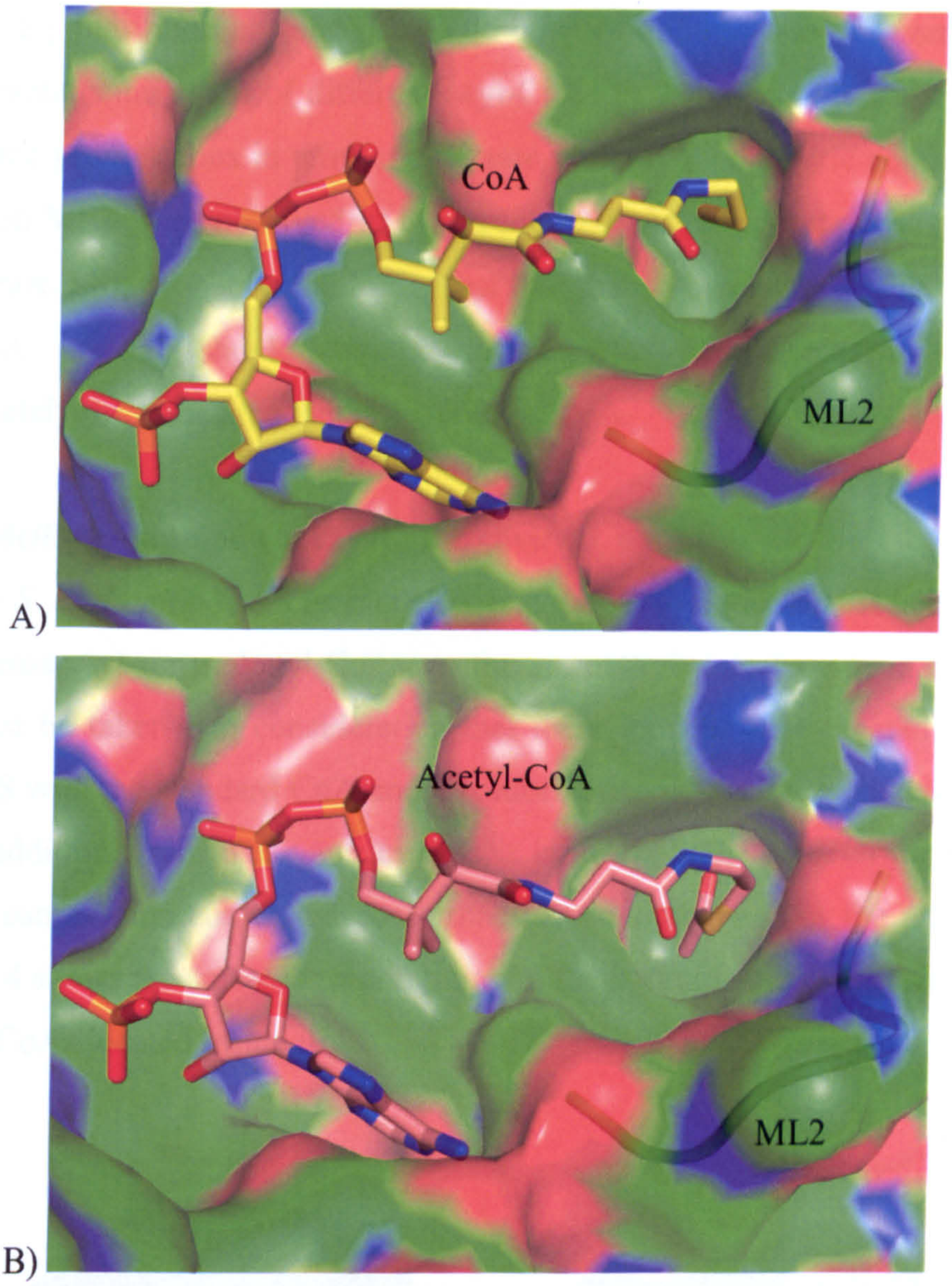


Figure 5.13 Charged surface representation of the hydrophobic cavity and ML2 surrounding CoA (A) and acetyl-CoA (B). The neutral regions are coloured in green. The positive and negative areas are shown in blue and red respectively. CoA and acetyl-CoA are shown as sticks and are coloured in yellow and pink respectively. ML2 is shown as ribbon.

5.7.2 Role of mobile loop ML2

By comparison with the *B. subtilis*, *S. pneumoniae* and human AcpSs, the equivalent areas of ML2 are located in different positions (Figure 4.19 and 4.21). ML2 in the *S. coelicolor* AcpS is in a more open conformation, while in the *S. pneumoniae* AcpS, in complex just with 3',5'-ADP, ML2 is in a much closed position. The equivalent ML2 region in the human AcpS assumes a sort of intermediate conformation between the previous ones. It is possible to suggest that

the relatively high mobility of ML2 could be very important to accommodate the phosphopantetheine chain of different acyl-CoAs by enlarging the cavity, making this enzyme promiscuous in term of substrate variety. No dramatic conformational changes on ML2 were observed in the structure of AcpS in complex with acetyl-CoA (Figure 5.8), probably due to the fact that this cofactor is just slightly longer than CoA and it should assume a contracted conformation, occupying approximately the same area that is covered by CoA.

5.7.3 Modelling modified CoAs into the cavity

Using the *S. coelicolor* AcpS as model, it is possible to notice that it is difficult to fit the acetoacetyl- or malonyl-CoAs, or the aromatic benzoyl-CoA, in the cavity, but it must be taken into consideration that these models were created using the holo-AcpS with the standard CoA and a conformation of ML2 that is characteristic of it. In addition, from these pdb models it is not possible to determine if these cofactors can assume contracted conformations as it happens for the acetyl-CoA. Figure 5.14 shows the surface representation of the hydrophobic cavity with several modified CoAs located inside it.

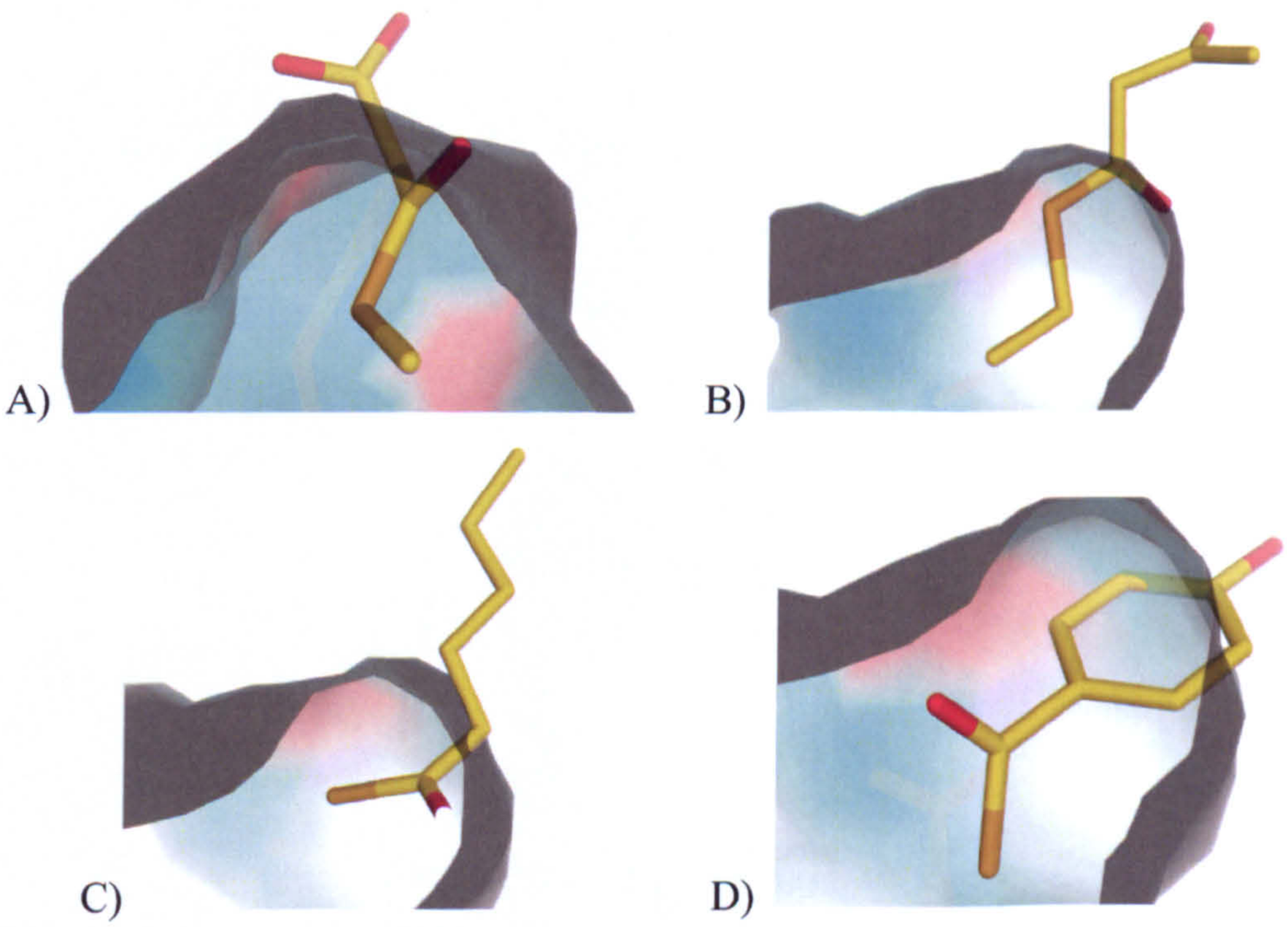


Figure 5.14 Modelling extended CoAs inside the hydrophobic cavity. A) malonyl group; B) acetoacetyl group; C) haxanoyl group; D) benzoyl group. AcpS subunit B is coloured in cyan, negative regions in red.

The phosphopantetheine chains of these modified CoAs were superimposed to the tail of CoA in the holo-AcpS structure, and it was tried to find the best possible conformations that allow the least steric clash between the phosphopantetheine chains and the protein. The cavity is well defined and seems to allow space for cofactors longer than CoA or acetyl-CoA. However, in this model the pocket is too small for even the smaller substrate analogues, malonyl- or benzoyl-CoAs, suggesting that this loop should assume a more open conformation *in vivo* to allow the binding of longer CoAs.

Chapter 6

Insight into AcpS mechanism and its interaction with ACP

In this chapter, a model for the interaction between the *S. coelicolor* AcpS and the *S. coelicolor* ACP, and for the pantetheinylation reaction mechanism, are presented. The proposed models are not based on a crystal structure of the complex between AcpS and ACP, but on the analysis and the comparison of several pdb structures. So, in order to clarify the role of these two proteins and to obtain a clearer picture of the reaction mechanism, efforts were also made to try to purify and crystallize this complex *in vitro*.

6.1 Interaction between AcpS and ACP

6.1.1 Proposed ACP recognition and interaction

After ACP binding, phosphopantetheine transfer occurs. For the type I fatty acid *B. subtilis* AcpS-ACP complex (PDB code=1f80), it was proposed that the contacts between the two proteins are predominantly hydrophilic with a determinant interaction between aspartate and glutamate residues on ACP helix α3 and arginine residues on AcpS helix α1 (Figure 6.1) [86]. This interaction places the ACP at the bottom of the active site orienting the activating ACP serine correctly. On the contrary, for the type II proteins, although the positioning of ACP near the active site is the same, hydrophobic interactions seem to contribute more strongly to the overall binding [66].

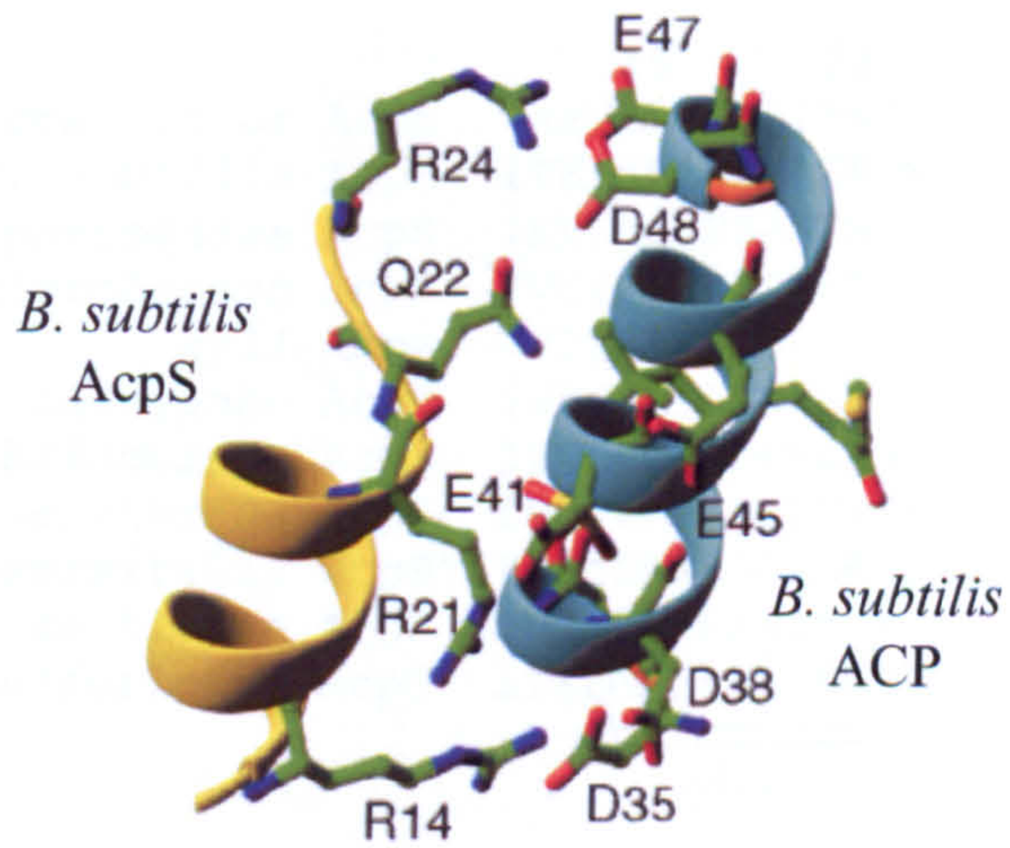


Figure 6.1 *B. subtilis* AcpS-ACP interaction. AcpS arginine residues can interact with aspartate and glutamate residues on the ACP (Figure adapted from [86]).

AcpS arginines and ACP negatively charged residues are well conserved among different homologous proteins (Figures 6.2 and 6.3). *S. coelicolor* Arg15 is conserved among all the other type I AcpSs, with the only exception of *S. pneumoniae* AcpS, while Arg22 and Arg29 are conserved in 9 and 10 proteins respectively. In the *B. halodurans* AcpS, both arginines are substituted by positive lysine residues, and in the *T. denticola* AcpS another lysine is very close to the position of the conserved Arg22. With regard to negative residues on ACP, Asp41 and Glu47 are completely conserved, while Glu53 is present in 11 different ACPs. Where Glu53 is not conserved, other glutamates or aspartates are in close proximity. Asp32 and Asp37 are also quite well conserved among different ACPs.

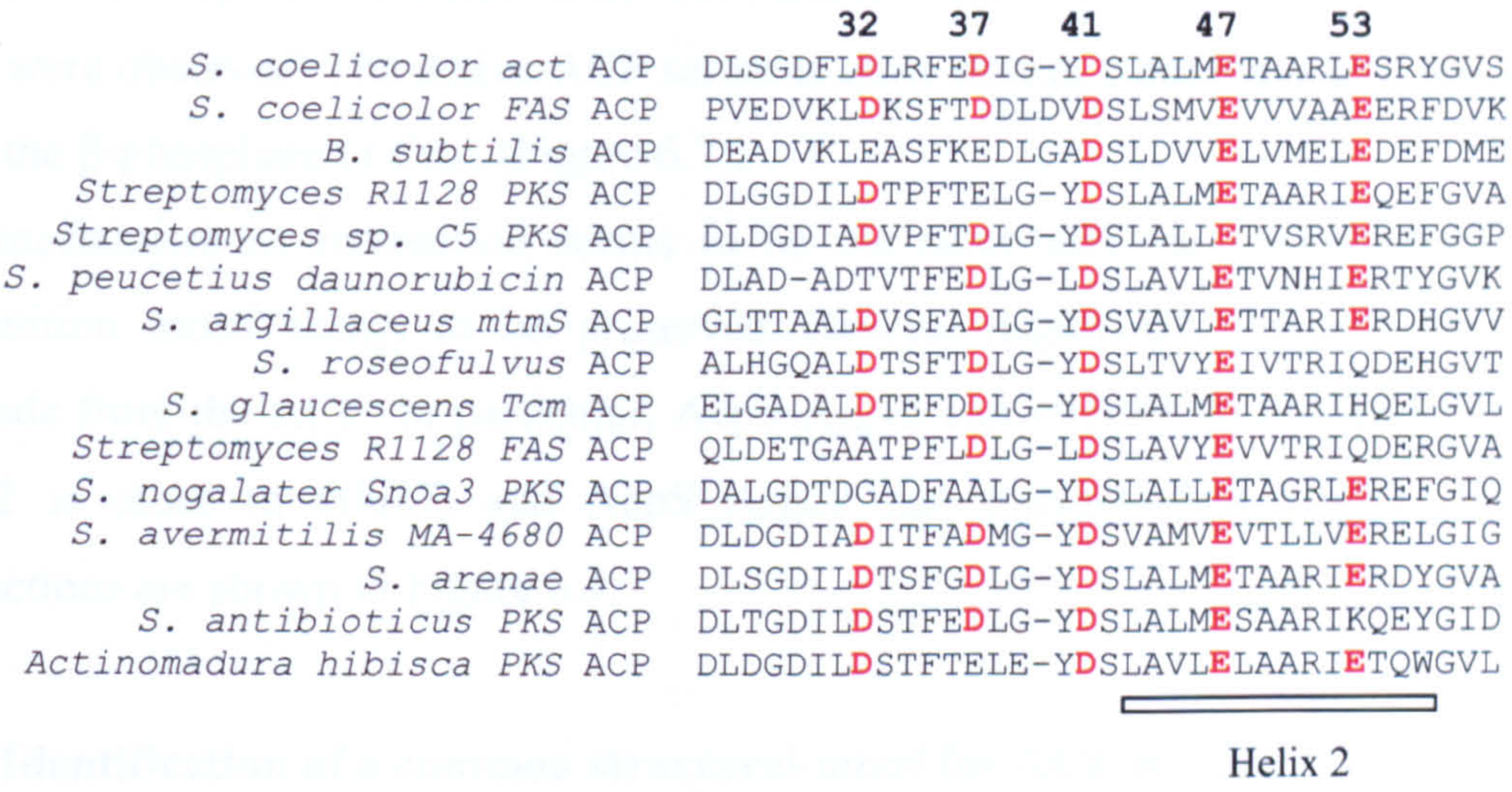


Figure 6.2 Multiple sequence alignment of polyketide and fatty acid ACPs. The conserved aspartate and glutamate residues are in bold and red.

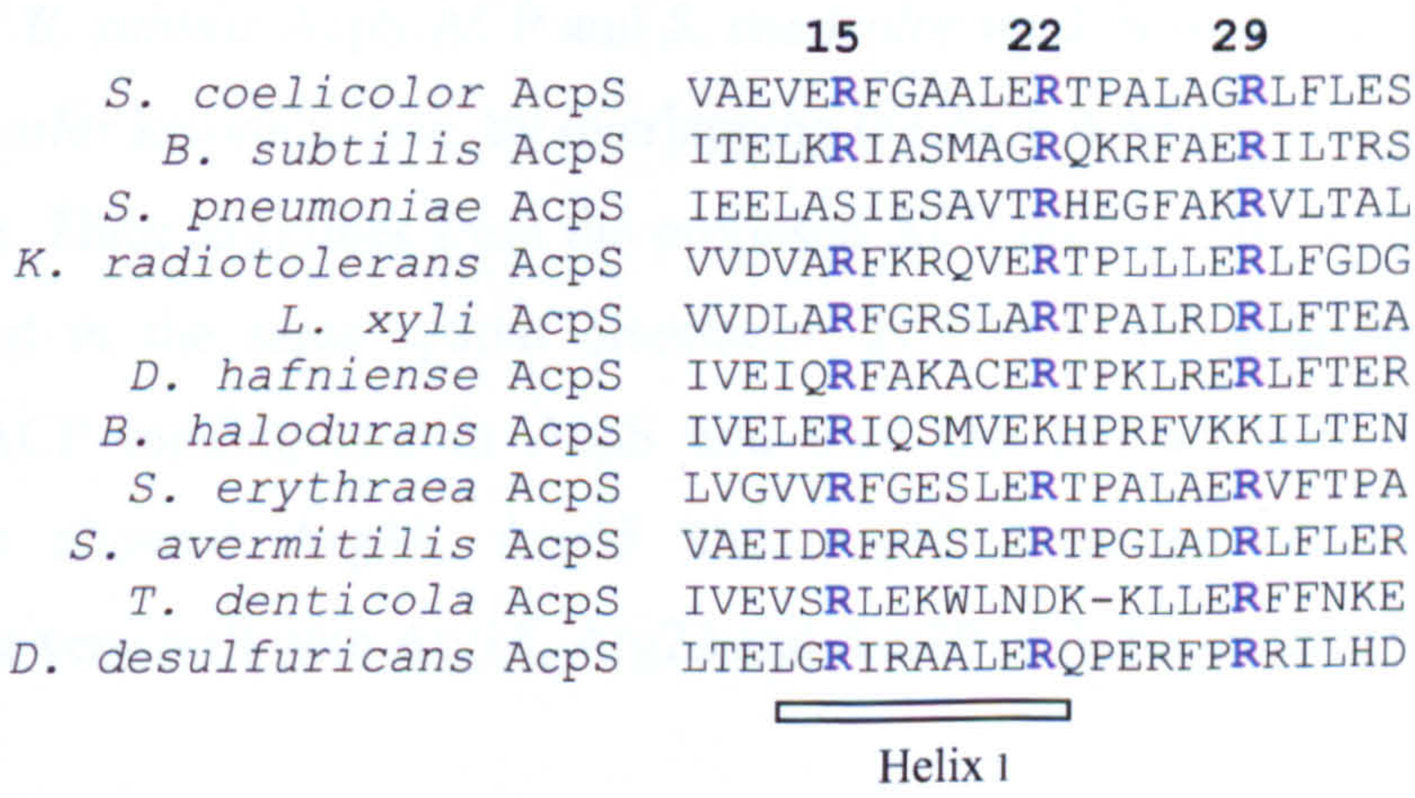


Figure 6.3 Multiple sequence alignment of polyketide and fatty acid AcpSs. The conserved arginines are in bold and blue.

The structure of the *B. subtilis* AcpS and ACP complex reveals a trimeric arrangement with a central core of AcpS molecules binding ACP on the outside with a stoichiometry of 1:1 [86]. Since this is the only AcpS:ACP complex structure available for type I AcpSs, it was utilized as a model for analysing the interaction with ACP and the reaction mechanism.

To investigate the *S. coelicolor* ACP binding to the AcpS, an AcpS:ACP complex model was made by superimposing the *S. coelicolor* holo-AcpS trimer to the *B. subtilis* AcpS, and a *S. coelicolor* apo-ACP NMR structure, kindly donated by Dr Matt Crump, on *B. subtilis* ACP, as described in chapter 3.8.7. In this model, AcpS helix α1 is very close to ACP helix α2. CoA is located at the interface between one monomer of AcpS and a molecule of ACP, and no clashes between the cofactor and ACP were observed. The active ACP serine is quite distant from the active site and from the β-phosphate of CoA (Figure 6.7).

The mechanism of interaction seems to be the same and the positive-negative recognition motif seems to be preserved. Several negatively charged residues protrude from the ACP. In particular, AcpS Arg15 can interact with Asp41, AcpS Arg22 is close to Glu47, and AcpS Arg29 can interact with Glu53. These interactions are shown in Figure 6.4.

6.1.2 Identification of a common structural motif for ACP binding

A set of arginine residues were also observed and proposed as a potential ACP binding site in the *S. coelicolor* ketoreductase [103], one of the enzymes *S. coelicolor* ACP interacts with during the synthesis of polyketides. The known complex of *B. subtilis* AcpS-ACP and *S. coelicolor* models were superimposed on the *S. coelicolor* ketoreductase, by overlapping the ACP modelled into that putative binding site. Three arginines from the proposed ACP binding site in ketoreductase are disposed in the same spatial orientation as that of the arginines from the proposed ACP binding site in AcpS and they can interact with ACP. Small adjustments allowed Arg38, Arg65 and Arg93 from the ketoreductase to superimpose very well with Arg15, Arg22 and Arg29 of AcpS (Figure 6.5).

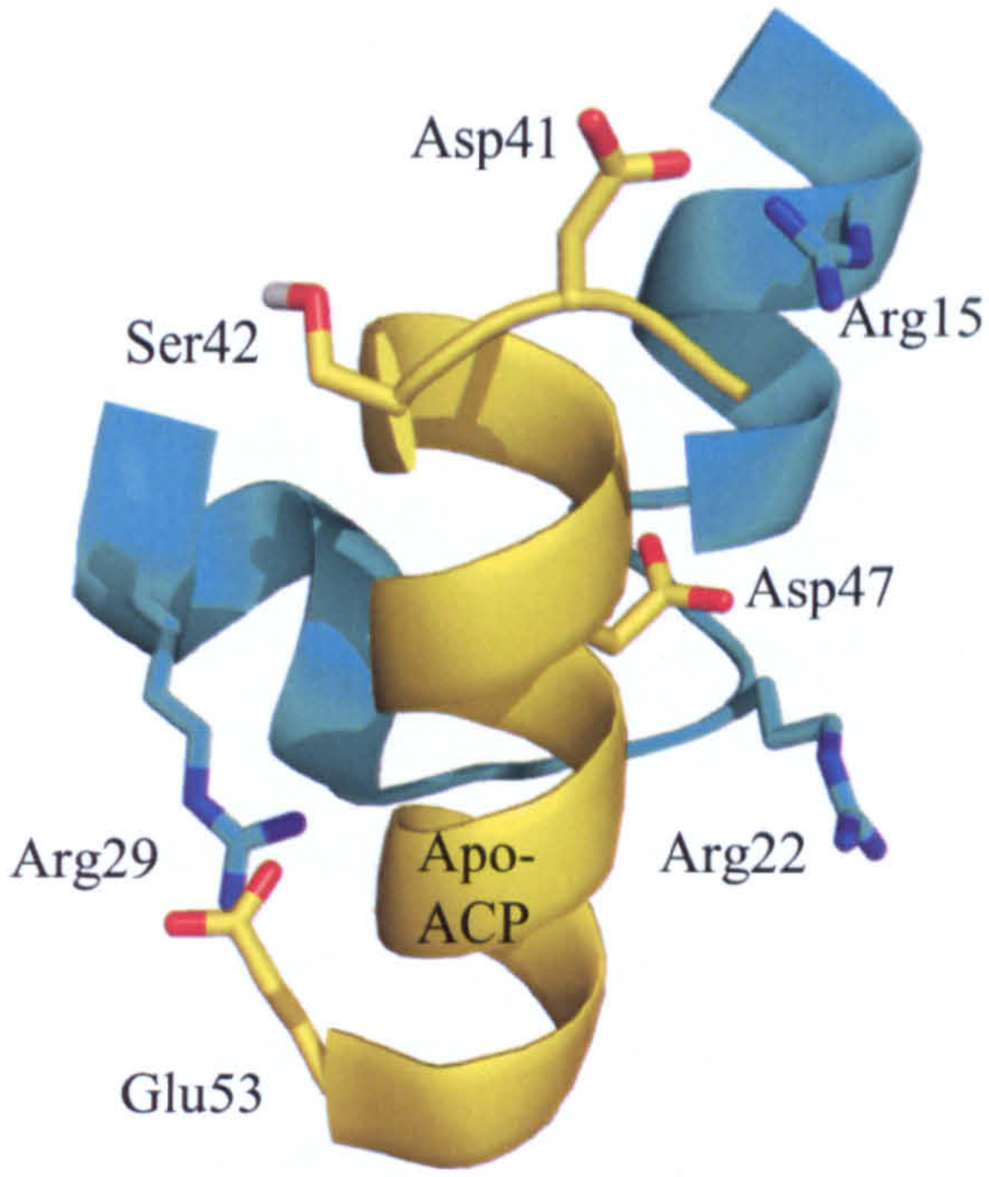


Figure 6.4 Model of *S. coelicolor* ACP interacting with AcpS. Cartoon representation of AcpS helix α1 and ACP helix α2. AcpS is coloured in cyan, ACP in yellow. The side chains of the residues involved in the interaction are shown as sticks.

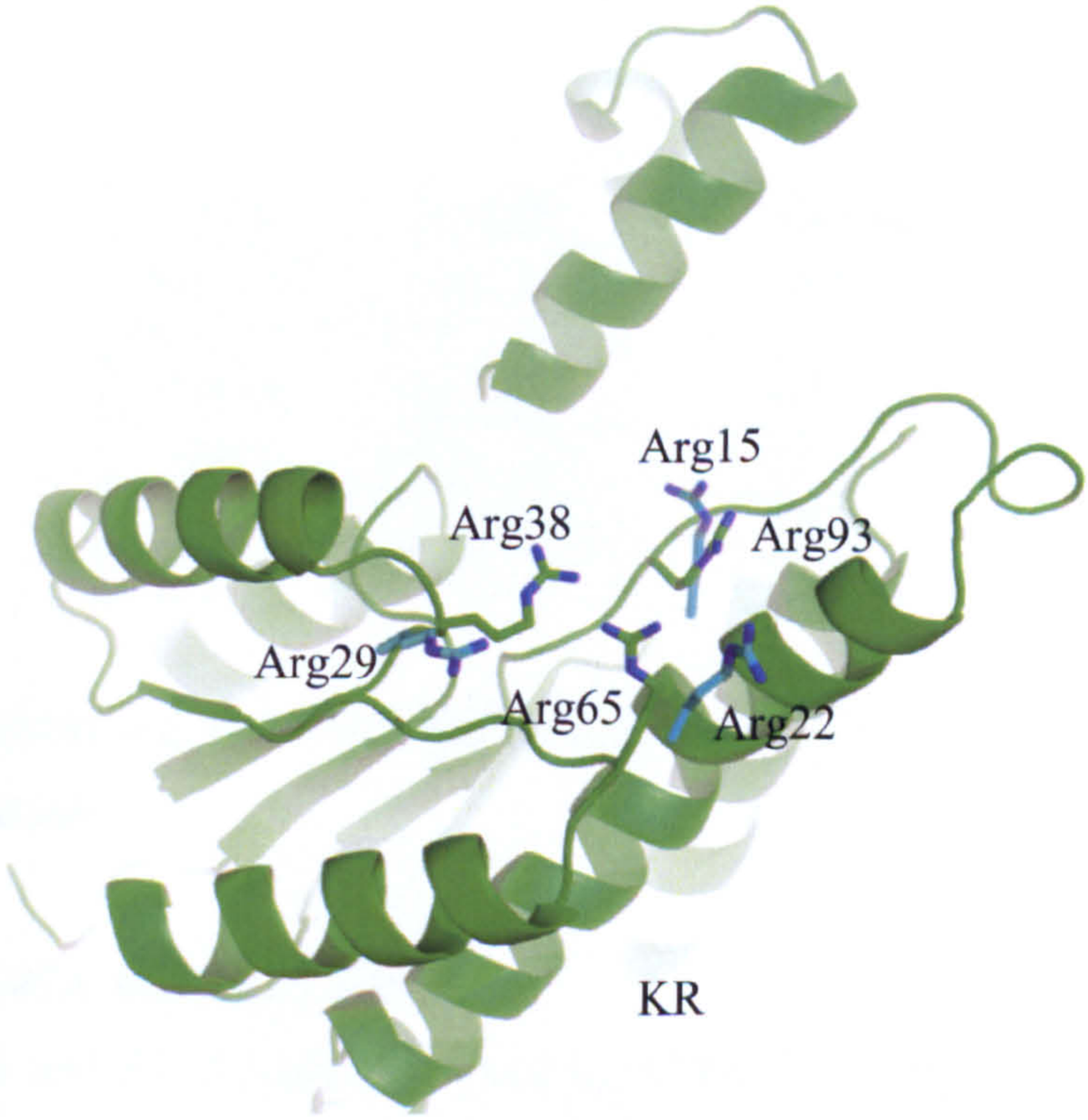


Figure 6.5 Proposed ACP binding site for KR. Cartoon representation of the proposed ACP binding site in ketoreductase (coloured in green). The side chains of the arginines from the AcpS are shown in cyan. These three AcpS arginines overlap very well with three KR arginines.

6.1.3 Analysis of mutational ACP data

The pdb model for the *S. coelicolor* ACP and AcpS interaction and binding was very useful to analyze mutational data on the ACP that were carried out by Dr Chris Arthur in the Department of Chemistry. Several C17S ACP double mutants were expressed and purified: C17S D41A, C17S D41E, C17S E47A and C17S E53A. The mutated residues, conserved among other fatty acid or polyketide ACPs, were supposed to be involved in the interaction with the AcpS, as proposed for the *B. subtilis* complex (Figure 6.4). Table 6.1 summarizes the results of the tryptophan fluorescence experiments and ESMS analysis of percentage of conversion to holo-ACP of these mutants. The affinity for apo-AcpS is within the same order of magnitude for each different ACP mutant, suggesting that, even though the selected ACP residues are important for the interaction with the AcpS and are necessary to tune the precise location of ACP in the AcpS active site to allow catalysis, other hydrophobic interactions are also involved and participate to lock the ACP in the right position. For example, from the model it is possible to notice that ACP Leu43 can extend into a pocket formed by AcpS Leu20, Leu26, Leu30, Phe16, Phe53 and Trp71.

ACP	K_d (μM)	% of conversion
Apo-ACP	1.11 ± 0.05	100%
Holo-ACP	2.07 ± 0.07	/
D41A	7.02 ± 0.21	0 %
D41E	1.42 ± 0.06	15 %
E47A	1.81 ± 0.22	100 %
E53A	2.07 ± 0.11	75 %

Table 6.1 Affinity for apo-AcpS and percentage of conversion to holo-ACP of C17S ACP mutants.

6.1.3.1 ACP E47A and E53A mutants

Mutants E47A and E53A both have similar affinity for complex formation as the product of the reaction, holo-ACP, which is about 2-fold reduced compared to the substrate, apo-ACP. Therefore, there is not much difference in affinity between the apo and the holo-protein. These mutations do not destroy the interaction between AcpS and ACP, suggesting that other interactions are involved. However, the

change in activity related to E53A ACP could be indicative of a role of Glu53 in the precise positioning of the loop where the ACP serine is located. While the mutation of Glu47 to an alanine does not affect the percentage of conversion to holo-ACP, the mutation of Glu53 to an alanine reduces its efficiency by 25%. Glu53, located at the top of the helix, is probably more important than Glu47 to lock the position of the ACP with respect to the AcpS and placing the loop containing the active serine close enough to the AcpS active site.

6.1.3.2 ACP D41A and D41E mutants

The mutagenesis of the conserved Asp41 into an alanine results in the only significant change to binding affinity, which is reduced seven fold compared to apo enzyme. In addition, this enzyme is completely inactive in phosphopantetheine transfer. The more conservative substitution of aspartate with glutamate, residue that has similar chemical properties but is just slightly longer, reduces the affinity for complex formation just slightly. The percentage of conversion to holo-ACP is reduced by approximately seven fold for this mutant. This suggests that ACP Asp41 can have a specific role in promoting catalysis, and not only a role in binding to the AcpS. The pK_a of both aspartate and glutamate are extremely similar. By inserting the extra carbon atom, the position of ACP and therefore of the serine attachment site could be probably altered. So, even if Asp41 is not involved in any chemical reaction, it seems to be important for locating the region containing the activating ACP serine in the right position. Its substitution with a glutamate could change the orientation of ACP Ser42 within the active site.

6.2 Proposed reaction mechanism

6.2.1 Initiation of ACP binding

In the holo-AcpS, helix α1 and MH1 are displaced in comparison with the apo-protein. The movement of these portions can be a possible signal to ACP that CoA is bound and AcpS is ready to react (Figure 6.6).

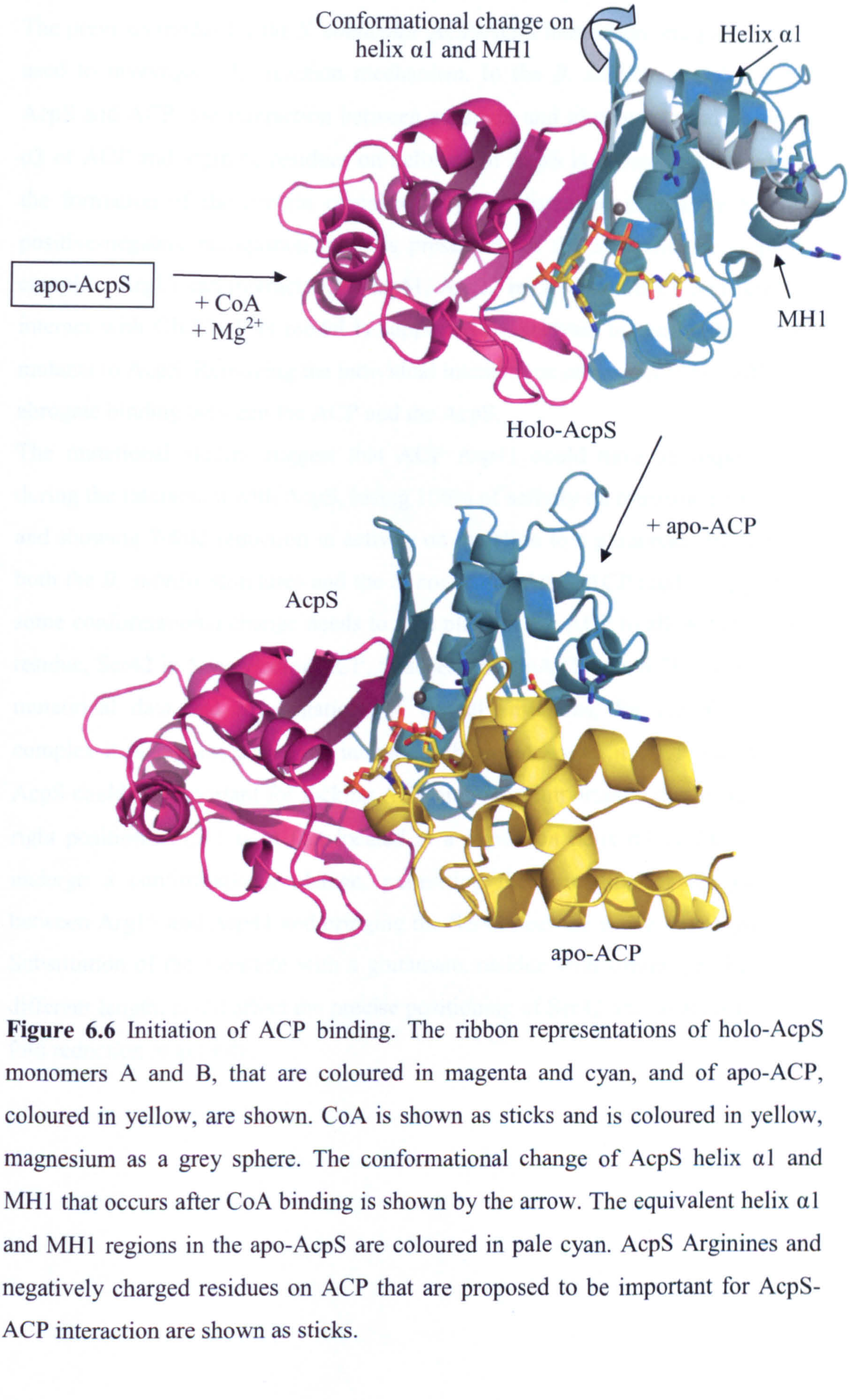


Figure 6.6 Initiation of ACP binding. The ribbon representations of holo-AcpS monomers A and B, that are coloured in magenta and cyan, and of apo-ACP, coloured in yellow, are shown. CoA is shown as sticks and is coloured in yellow, magnesium as a grey sphere. The conformational change of AcpS helix α1 and MH1 that occurs after CoA binding is shown by the arrow. The equivalent helix α1 and MH1 regions in the apo-AcpS are coloured in pale cyan. AcpS Arginines and negatively charged residues on ACP that are proposed to be important for AcpS-ACP interaction are shown as sticks.

6.2.2 *S. coelicolor* AcpS:ACP interaction and serine positioning

The previous model for the *S. coelicolor* ACP-AcpS interaction (chapter 3.8.7) was used to investigate the reaction mechanism. In the *B. subtilis* complex between AcpS and ACP, the interaction between aspartate and glutamate residues on helix α3 of ACP and arginine residues on helix α1 of AcpS is considered important for the formation of the protein complex [86]. As discussed in chapter 6.1.1, this positive-negative recognition motif is preserved for the *S. coelicolor* AcpS-ACP complex. Arg15 can interact with Asp41, Arg22 is close to Glu47, and Arg29 can interact with Glu53. This model is supported by analysis of the binding of ACP mutants to AcpS. Removing the individual interactions seems not to be sufficient to abrogate binding between the ACP and the AcpS.

The mutational studies suggest that ACP Asp41 could have an important role during the interaction with AcpS, losing 100% of activity on mutation to an alanine and showing 7-fold reduction in activity on mutation to a glutamate. Inspection of both the *B. subtilis* structures and the *S. coelicolor* AcpS-ACP model suggests that some conformational change needs to take place in the ACP to allow the modified residue, Ser42 in *S. coelicolor* ACP, to access the CoA (Figure 6.7). So, after these mutational data and investigations of crystal structures, for the *S. coelicolor* complex it is suggested that the interaction between Asp41 in ACP and Arg15 in AcpS could be important for locking ACP and more importantly the Ser42 into the right position. Asp41 is in fact located in a region on helix α2 of ACP that can undergo a conformational change, permitting the formation of the salt bond between Arg15 and Asp41 and bringing the Ser42 near the active site (Figure 6.8). Substitution of the aspartate with a glutamate, residue with similar pK_a but with a different length, could affect the precise positioning of Ser42 and so account for a 7 fold reduction in activity.

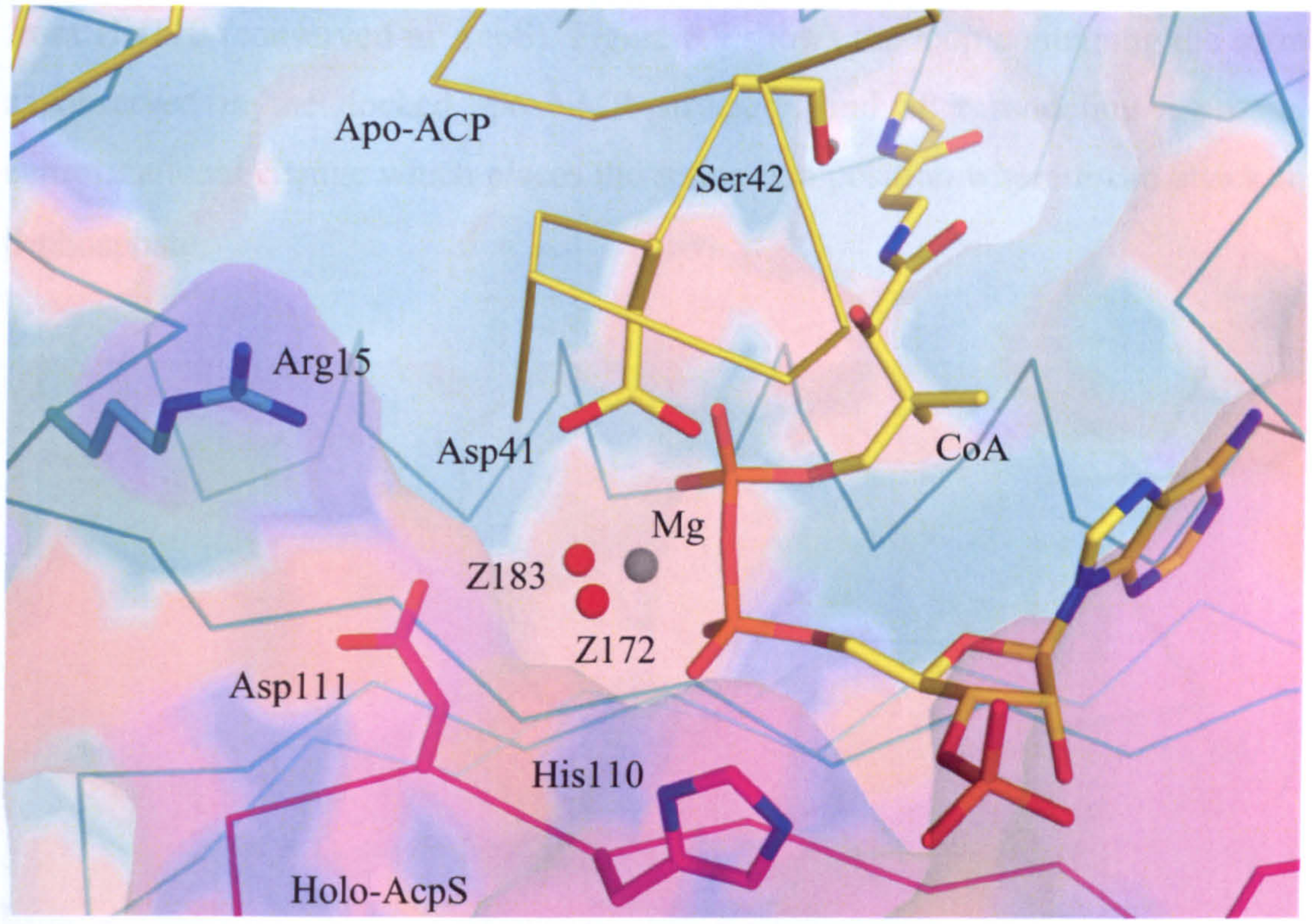


Figure 6.7 AcpS:ACP active site. The figure shows the beginning of the reaction: the interaction between AcpS Arg15 and ACP Asp41 could be important to lock ACP and the loop containing the serine into the active site and in the right position. The charged surface and the ribbon representation of AcpS monomers A and B are shown. The neutral regions are coloured in magenta (momomer A) and cyan (monomer B), while the positive and negative areas are shown in blue and red respectively. The ribbon representation of apo-ACP is coloured in yellow. The side chains of residues proposed to be involved in catalysis are shown as sticks. Water molecules are coloured in red, magnesium in grey.

For the *B. subtilis* AcpS, a water molecule was proposed to deprotonate the ACP serine [86], while for the Sfp protein it was suggested that the hydroxyl group of the active serine could take the place of a water molecule and be deprotonated by a base [78]. With small adjustments to the loop on which Asp41 and Ser42 are located, it is possible to place the *S. coelicolor* Ser42 in a position where it could attack the β-phosphate from either side (Figure 6.8), for example replacing water Z183, which coordinates with Mg^{2+}, and is also within hydrogen bonding distance from Asp111, or water Z172 which is potentially within hydrogen bonding distance

from His110 (conserved in AcpS). Figure 6.9 shows the loop containing the serine as observed in the docked apo NMR structure, and after modeling a possible conformational change which places the serine in a position where it can attack the β-phosphate.

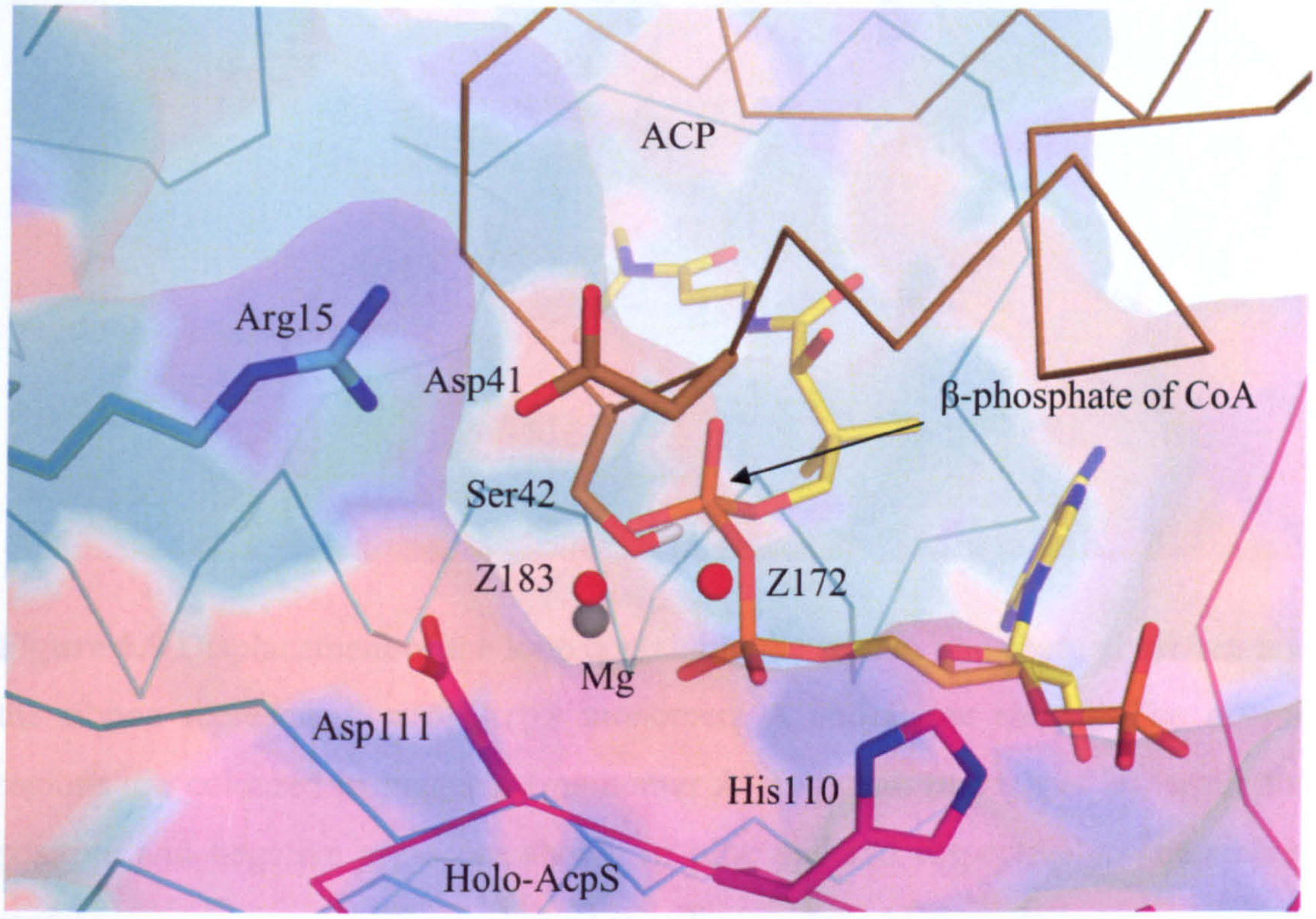


Figure 6.8 Positioning of Ser42 in the AcpS active site. The charged surface and the ribbon representation of AcpS monomers A and B are shown. The neutral regions are coloured in magenta (momomer A) and cyan (monomer B), while the positive and negative areas are shown in blue and red respectively. The ribbon representation of ACP is coloured in sand. The side chains of residues proposed to be involved in catalysis are shown as sticks. Water molecules are coloured in red, magnesium in grey

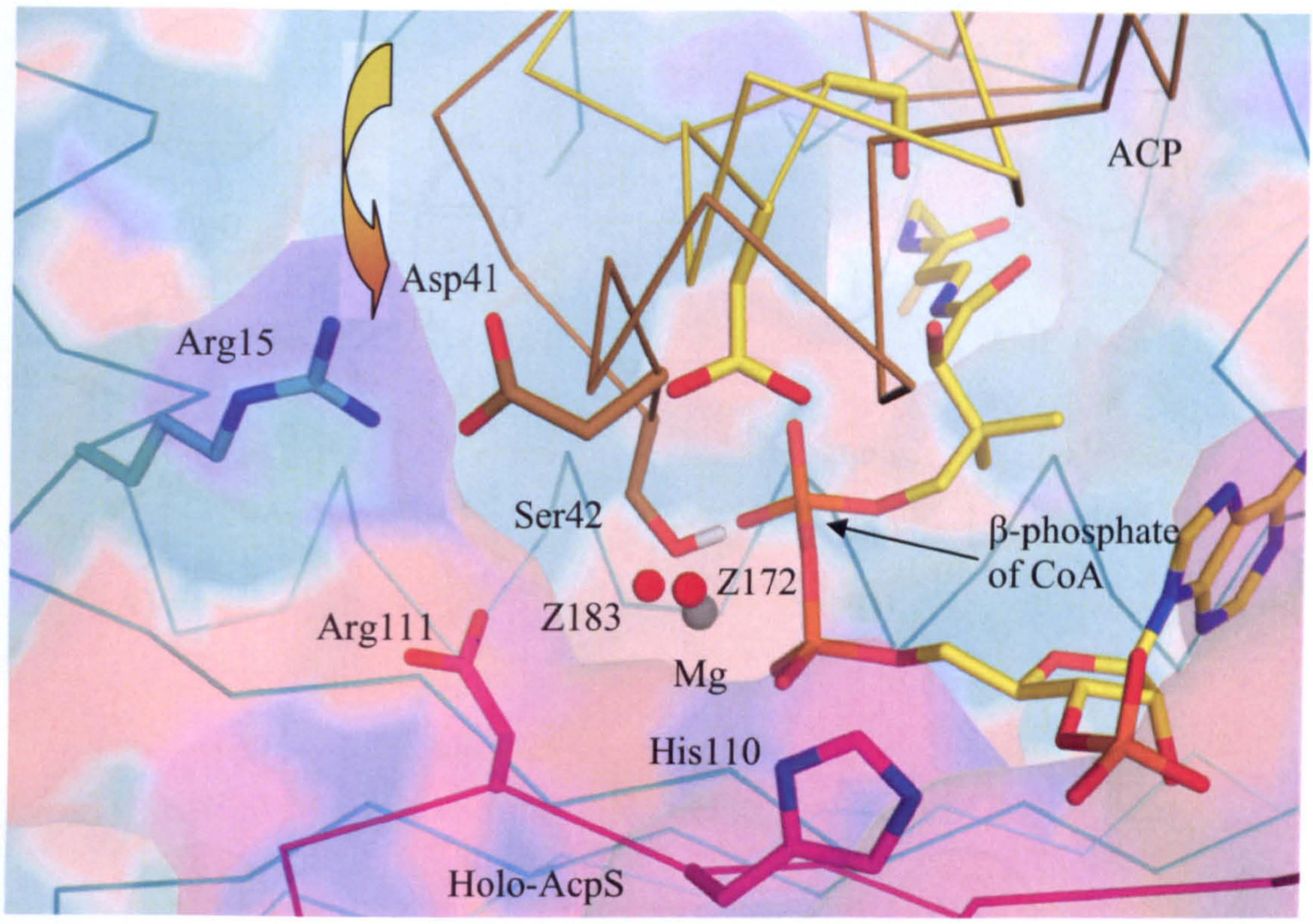


Figure 6.9 Displacement of the loop containing the serine. The charged surface and the ribbon representation of AcpS monomers A and B are shown. The neutral regions are coloured in magenta (momomer A) and cyan (monomer B), while the positive and negative areas are shown in blue and red respectively. The ribbon representation of apo-ACP is coloured in yellow, while ACP modelled in a position to attack the β-phosphate in the active site in sand. The side chains of residues proposed to be involved during catalysis are shown as sticks. Water molecules are coloured in red, magnesium in grey.

6.2.3 FAS *B. subtilis* AcpS proposed mechanism

For the *B. subtilis* type I fatty acid reaction [86], it was proposed that, after CoA and ACP binding, the phosphopantetheinylation of the ACP occurs. The ζ-amino group of Asp8 is deprotonated, while Lys62 is protonated. The hydrogen on Asp8 can be transferred to Lys62, as they are spatially adjacent. A water molecule in the AcpS active site can be activated by the removal of a proton, reaction that can be carried out by another water molecule or by ACP Asp35, which is spatially adjacent. This "activating water" can then deprotonate the hydroxyl group of the ACP serine, that can be placed near this water molecule. ACP Ser36 will carry out the nucleophilic attack on the β-phosphate of CoA (Figure 6.10).

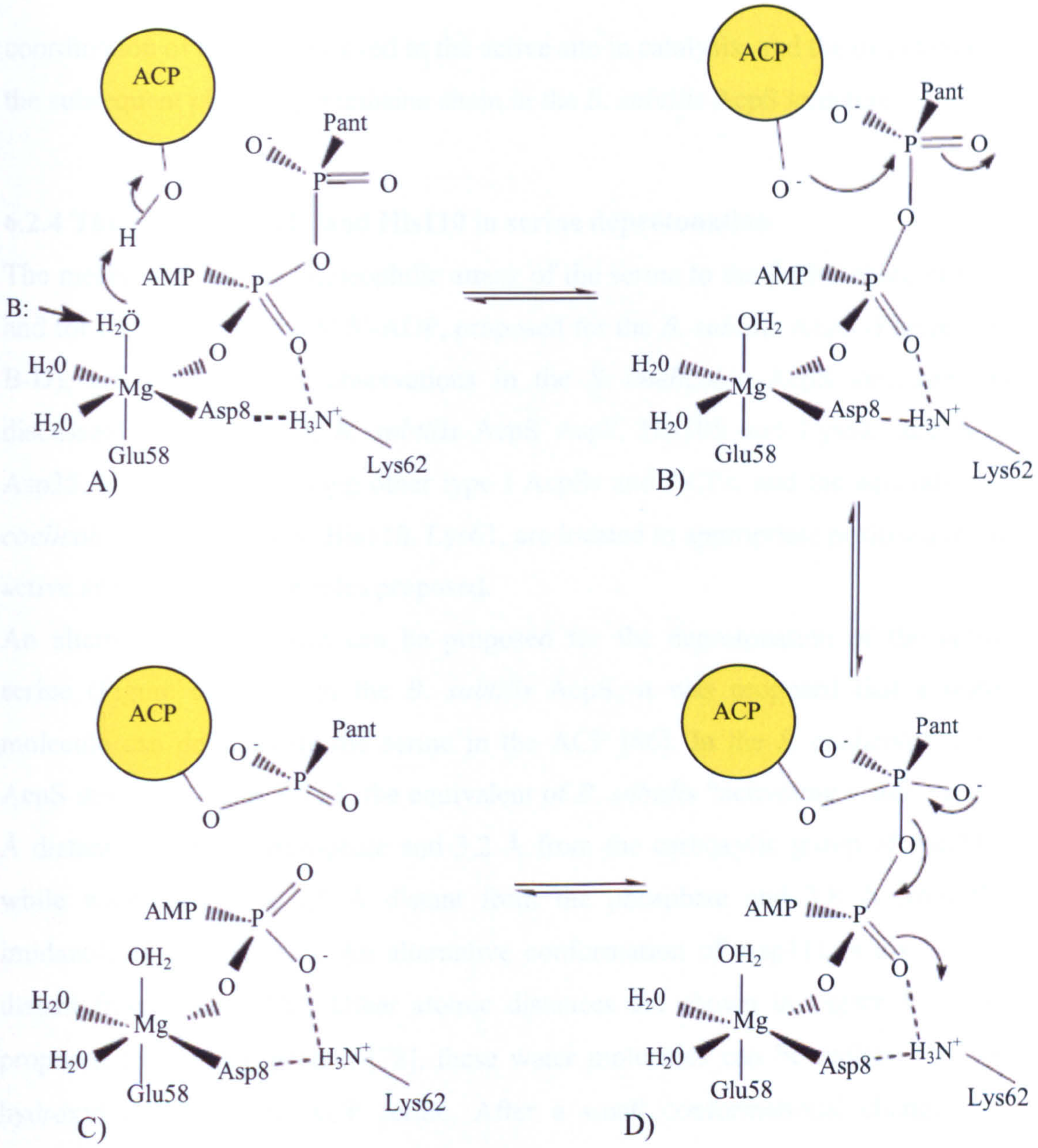


Figure 6.10 Reaction mechanism proposed for the *B. subtilis* fatty acid complex (Figure adapted from [86]). B: base; AMP: adenine monophosphate. Pant: phosphopantetheine chain.

Finally, after the transfer of the phosphopantetheine chain to Ser42, the 3',5'-ADP product can be stabilize by magnesium, Lys62 and the main chain amide of His105, and then after proton transfer from the solvent or Lys62 it can be released in solution. However, as described in chapter 4.8, this model is based on a crystal structure where a calcium and not magnesium is present in the active site, and CoA is in a different conformation in comparison to the *S. coelicolor* structure. The calcium can affect the conformation of the alpha and beta phosphates, the

coordination of residue involved in the active site in catalysis, and the orientation of the subsequent phosphopantetheine chain in the *B. subtilis* AcpS structure.

6.2.4 The role of Asp111 and His110 in serine deprotonation

The mechanism for the nucleophilic attack of the serine to the β-phosphate of CoA and for the release of the 3',5'-ADP, proposed for the *B. subtilis* AcpS (Figure 6.10 B-D), are supported by observations in the *S. coelicolor* AcpS structure. As discussed in chapter 4.7, *B. subtilis* AcpS Asp8, His105 and Lys62, and ACP Asp35, are conserved among other type I AcpSs and ACPs, and the equivalent *S. coelicolor* residues, Asp9, His110, Lys61, are located in appropriate positions in the active site to perform the roles proposed.

An alternative mechanism can be proposed for the deprotonation of the active serine (Figure 6.10 A). In the *B. subtilis* AcpS, it was proposed that a water molecule can deprotonate the serine in the ACP [86]. In the *S. coelicolor* holo-AcpS structure, water W183, the equivalent of *B. subtilis* "activating water", is 3.3 Å distant from the β-phosphate and 3.2 Å from the carboxylic group of Asp111, while water W172 is 4.5 Å distant from the phosphate and 3.8 Å from the imidazole ring of His110. An alternative conformation of Asp111 is also 1.6 Å distant from water Z183. Other atomic distances are shown in Figure 6.11. As proposed for the Sfp protein [78], these water molecules can be replaced by the hydroxyl group of the ACP serine. After a small conformational change, the aspartate or the histidine residues can deprotonate the hydroxyl group of Ser42 (Figure 6.11).

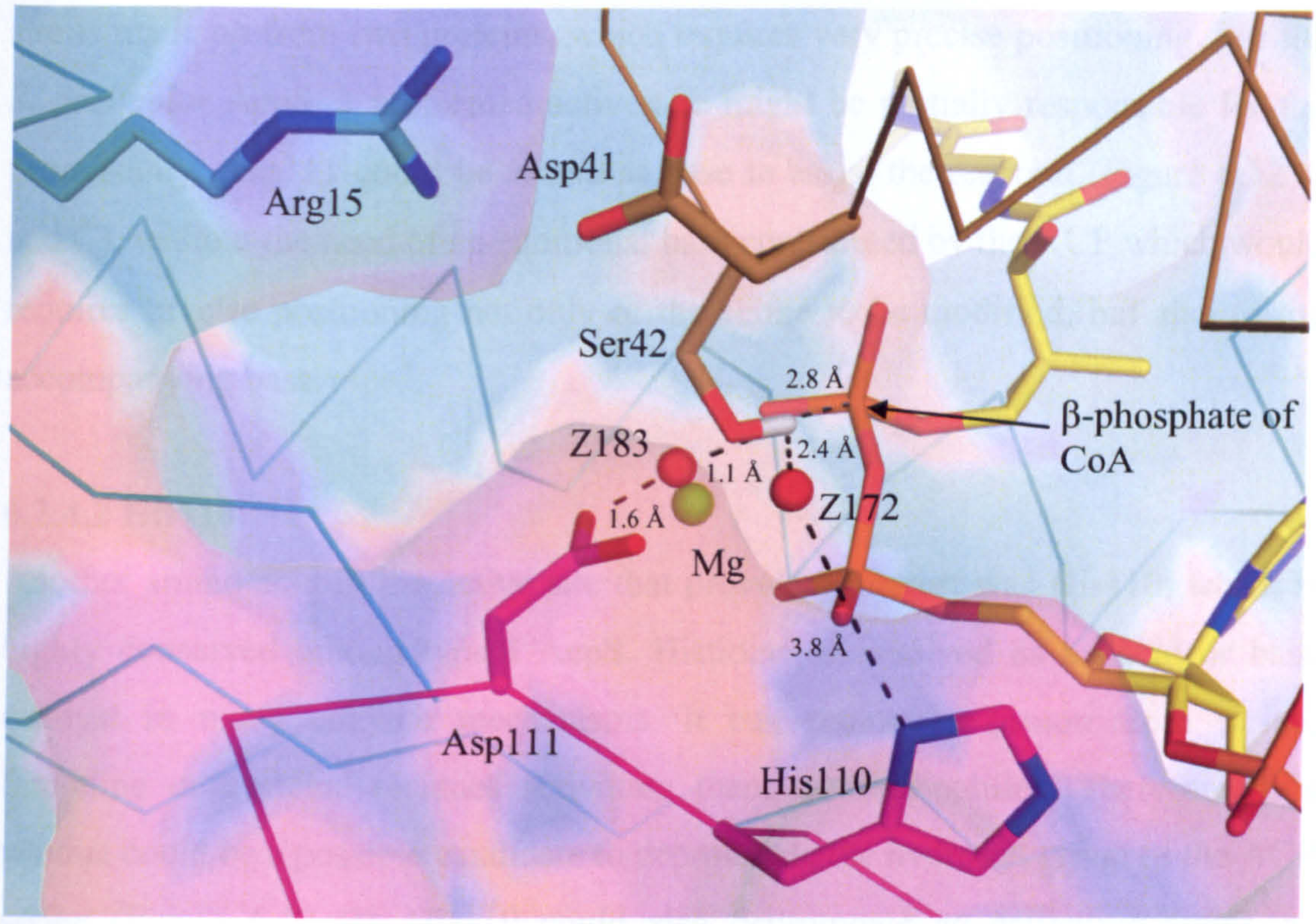


Figure 6.11 Possible catalytic role for AcpS Asp111 and His110. The charged surface and the ribbon representation of AcpS monomers A and B are shown. The neutral regions are coloured in magenta (momomer A) and cyan (monomer B), while the positive and negative areas are shown in blue and red respectively. The ribbon representation of apo-ACP is coloured in yellow, while ACP modelled in a position to attack the β-phosphate in the active site in sand. The side chains of residues proposed to be involved during catalysis are shown as sticks. Water molecules are coloured in red, magnesium in grey.

6.2.4.1 Asp111

At the pH where the phosphopantetheinylation occurs, the carboxyl group of the aspartate side chain is deprotonated, and its highly electronegative oxygens can remove a proton from the hydroxyl group of the serine. Asp111 could act in a similar way to the *B. subtilis* ACP Asp35 [86], but allowing less specific binding to promote a broader specificity for ACP recognition. Asp111 is not so well conserved among the other AcpS and it is replaced by a threonine in the *B. subtilis* AcpS. Normally, it is expected that catalytic residues are conserved among all the members of the same family. However, *S. coelicolor* AcpS is more promiscuous in substrate recognition. For the *B. subtilis* AcpS, it has been proposed that the active

site is made up from two proteins, which requires very precise positioning. For the *S. coelicolor* AcpS, a preformed active site might be partially responsible for the promiscuity. Asp111 could be *in situ* as base to assist the reaction (Figure 6.12 A and C), without the need of an additional base contributed by the ACP which would requires precise positioning not only of the serine to be modified, but also of the accompanying base.

6.2.4.2 His110

Another amino acid in the active site that provoked interest was His110, which is highly conserved among type I AcpS. Histidine is involved as an acid or base catalyst in many enzyme mechanisms. It can remove a proton from serine, threonine or cysteine residues activating them as nucleophiles. Therefore, this residue could be a possible candidate to deprotonate the hydroxyl group of the ACP serine (Figure 6.12 B). His110 could also deprotonate an activating water, as proposed by Parris (Figure 6.12 D).

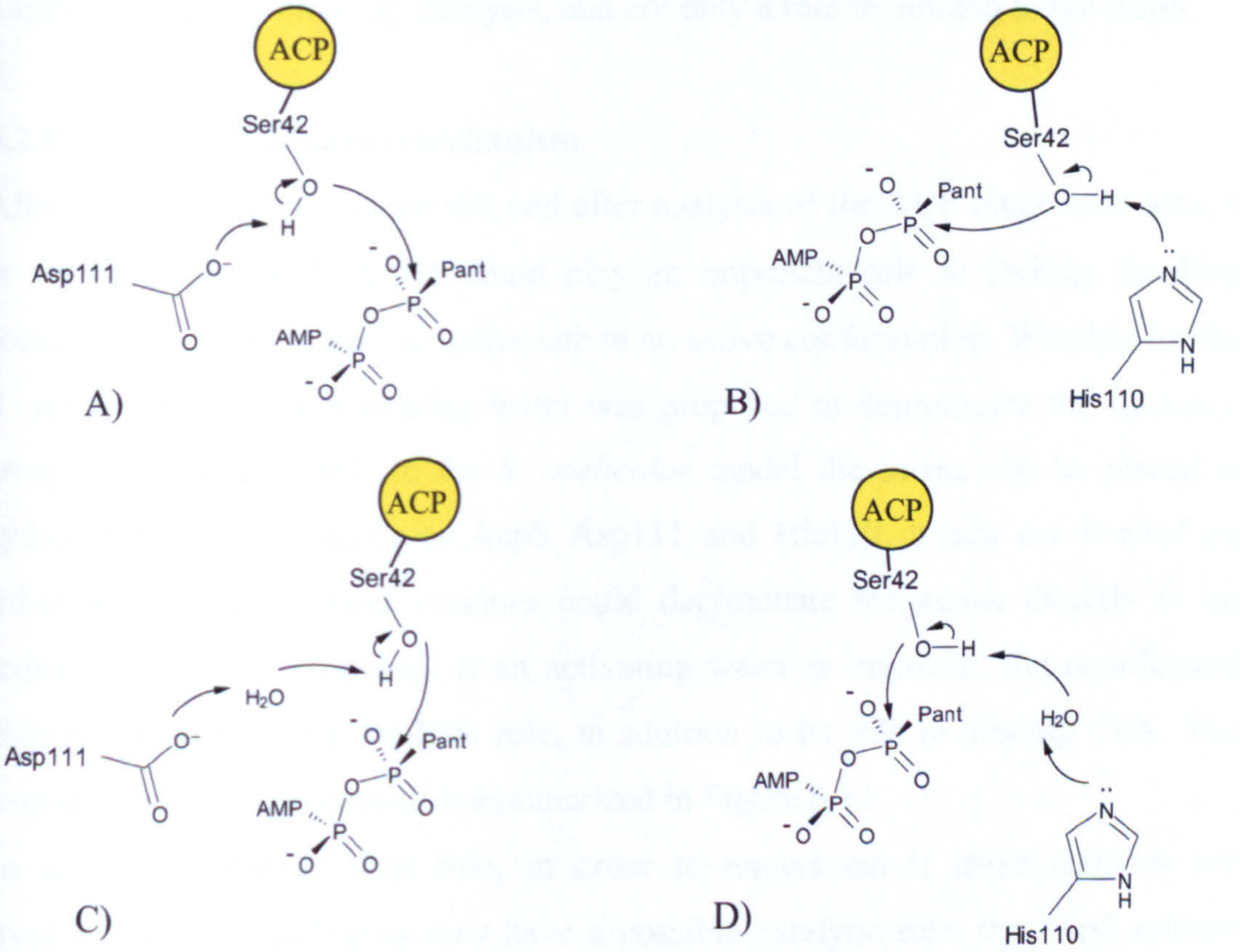


Figure 6.12 ACP serine deprotonation. AcpS Asp111 or His110 may deprotonate directly (A and B) or indirectly (via an activating mediator water, C and D) the active serine on ACP.

6.2.5 Conclusions

A model for the interaction with the ACP and for the reaction mechanism was proposed, based on the AcpS structures presented in this thesis and on a model of the *S. coelicolor* AcpS:ACP complex.

6.2.5.1 Interaction with ACP

Three conserved arginines on helix α1 of the AcpS can interact with aspartate and glutamate residues on helix α2 of ACP. Mutations of the ACP negatively charged residues gave further evidence for their role in AcpS binding. However, the affinity for apo-AcpS is within the same order of magnitude for each different ACP mutants. Even though the selected ACP residues are important for the interaction with the AcpS and are necessary for the precise positioning of ACP in the AcpS active site to allow catalysis, other hydrophobic interactions are probably involved and participate to lock the ACP in the right position. Mutation of the ACP Asp41 resulted to destroy the activity of the enzyme, suggesting that this residue may have a specific role in promoting catalysis, and not only a role in binding to the AcpS.

6.2.5.2 Proposed reaction mechanism

After inspection of the active site and after analysis of the ACP mutational data, it is proposed that ACP Asp41 could play an important role in locking the loop containing the Ser42 near the active site in an active conformation. Whether for the *B. subtilis* protein an activating water was proposed to deprotonate the hydroxyl group of the serine [86], in the *S. coelicolor* model the serine can be placed at hydrogen bonding distance to AcpS Asp111 and His110, which are located on either sides of CoA. Both residues could deprotonate the serine directly or an activating water as proposed. If an activating water is involved, the coordinated magnesium will have a catalytic role, in addition to its role in binding CoA. The proposed reaction mechanism is summarized in Figure 6.13.

To investigate further their role, in order to understand if these residues are involved in CoA binding or they have a possible catalytic role, the AcpS mutant H110A and D111A were cloned, crystallized and biochemically characterized (Chapter 7).

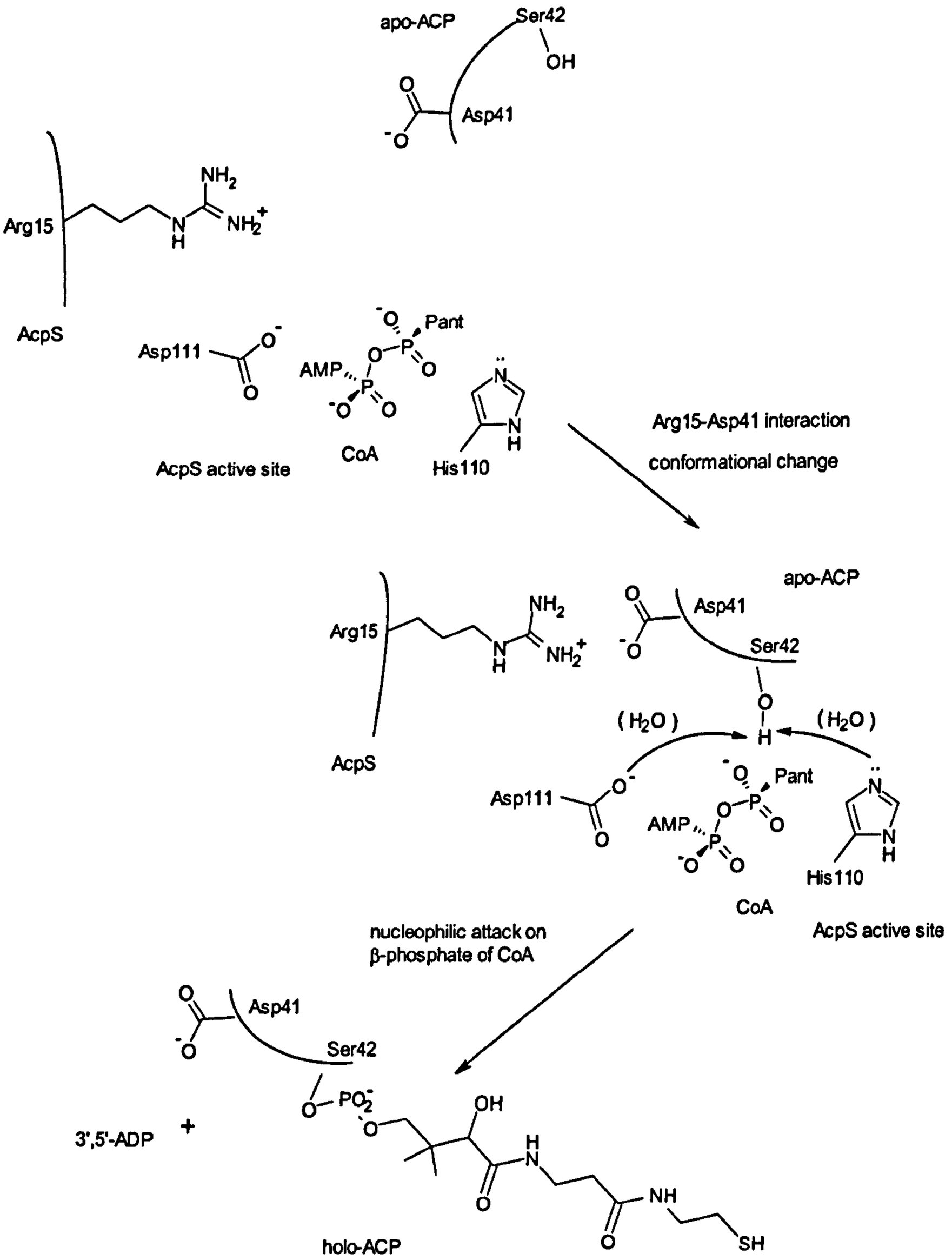


Figure 6.13 Proposed reaction mechanism. The interaction between ACP Asp41 and AcpS Arg15 could lock the loop containing the active serine in the right conformation allowing catalysis to occur. AcpS His110 or Asp111 could deprotonate (in)directly the serine, and after nucleophilic attack on the β-phosphate of CoA the phosphopantetheine chain is transferred to the ACP.

6.3 Purification trials of the AcpS-ACP complex

The model proposed for the interaction between ACP and AcpS and for the reaction mechanism was based only on the analysis of PDB structures. In order to understand better the reaction mechanism and to confirm the hypothesis made based on a model, the natural following step was to try to purify and crystallize this complex *in vitro*. So, ACP was firstly expressed and purified, and then purification and crystallization trials of the AcpS-ACP complex were carried out.

6.3.1 ACP expression

The *S. coelicolor* actinorhodin C17S ACP DNA cloned in the pET11c vector was kindly donated by Dr Matt Crump and Dr John Crosby (Department of Chemistry). As can be seen from the sequence of the protein shown in Figure 6.14, the protein is not tagged. Protparam [134] calculated a molecular weight of about 9200 Da, a pI of 3.98 and a extinction coefficient of 0.323. No tryptophanes are present, and the UV absorbance is determined by the presence of 2 phenylalanines and 2 tyrosines.

Prior to expression, the pET-11c-ACP DNA was transformed in fresh *E. coli* BL21(DE3) (chapter 3.4.5). The success of the transformation was checked by extracting plasmid from transformants (chapter 3.4.2), subjecting the DNA to endonuclease restriction digest (chapter 3.4.3) and checking the size of the fractions on an agarose gel. ACP was then expressed by induction with 1 mM IPTG at 37°C, and the expression continued for 4-5 hours at 32°C.

```
MATLLTTDDLRRALVESAGETDGTDLSGDFLDLRFEDIGYDSLALMETAARLESRY
GVSIPDDVAGRVDTPRELLDLINGALAEAA
```

Figure 6.14 Amino acid sequence of the actinorhodin C17S ACP.

6.3.2 ACP purification

The protein was not tagged, and after sonication, the supernatant was loaded into the anion exchange column (chapter 3.6.4). The chromatogram (Figure 6.15 A) presents a lot of peaks, and the small but quite defined one at about 360 mM NaCl corresponds to ACP. As shown on the gel (Figure 6.15 B), the level of ACP purity

in these fractions was not very high, so a further gel filtration step (chapter 3.6.5) was used to get rid of the contaminants. This step was very effective due to the small size of ACP in comparison with the other proteins present in solution. ACP was eluted as a well-defined single peak during gel filtration (Figure 6.16 A), and the SDS-PAGE gel loaded with the correspondent fractions (Figure 6.16 B) shows that it is very pure. Finally, the resulting apo-ACP was desalted (chapter 3.6.8) and stored at -80ºC until required. From previous experiments carried out in the Department of Chemistry, it was known that ACP can run on a SDS-PAGE gel with a lower apparent weight. The band corresponding to the ACP is indeed located below the 6.5 kDa band of the molecular markers.

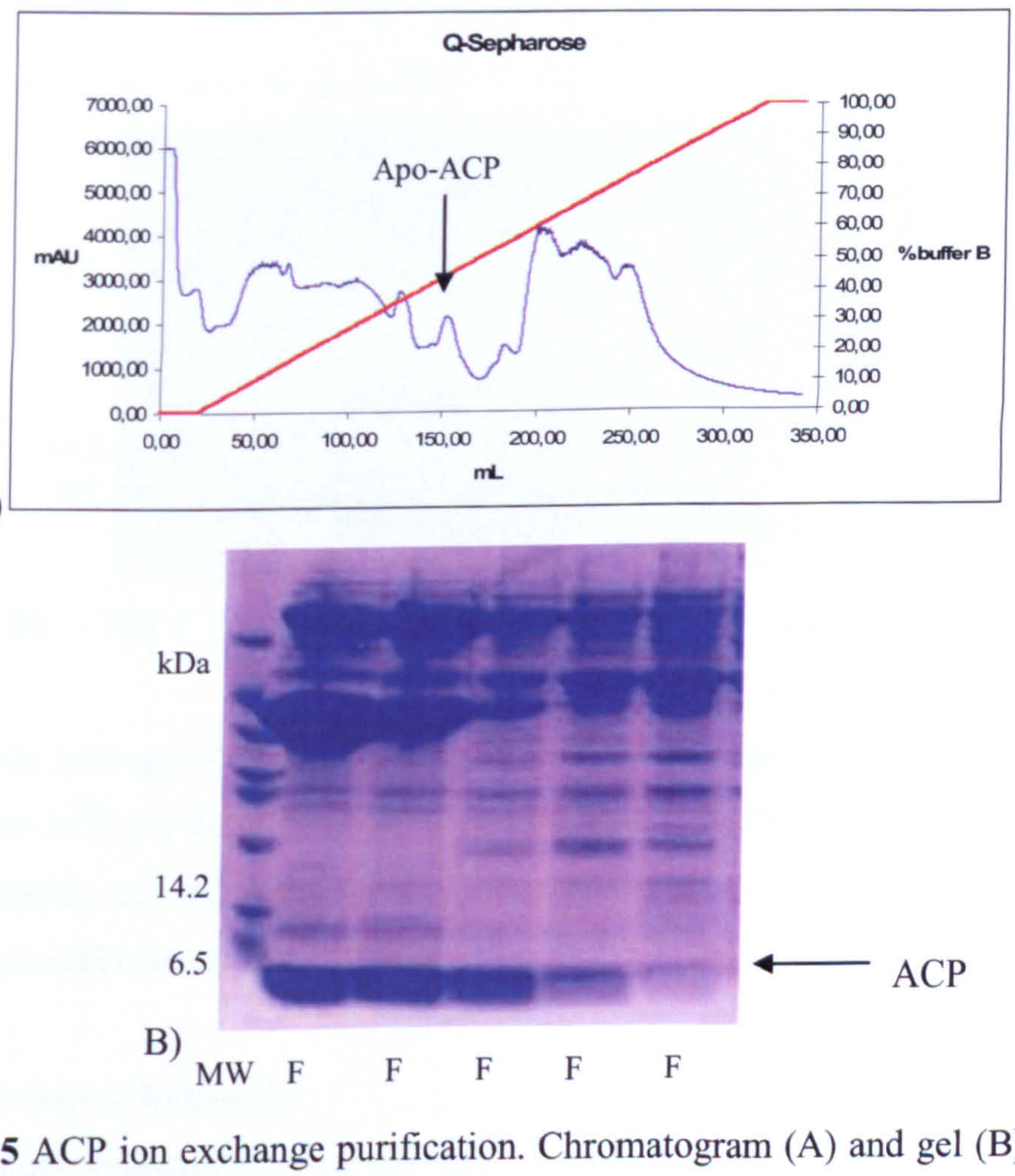


Figure 6.15 ACP ion exchange purification. Chromatogram (A) and gel (B) after the ion exchange column. A) The arrow indicates when ACP is eluted. UV absorbance is shown in blue, the gradient of buffer B in red. B) SDS-PAGE gel corresponding to the ACP peak. M: Molecular weights; F: fractions of apo-ACP from the related peak.

In addition, more than one band can be observed, probably due to protein oligomerization (Figure 6.16 B). The addition of DTT resulted in a decrease in the aggregation, and therefore in clearer bands on SDS-PAGE gels.

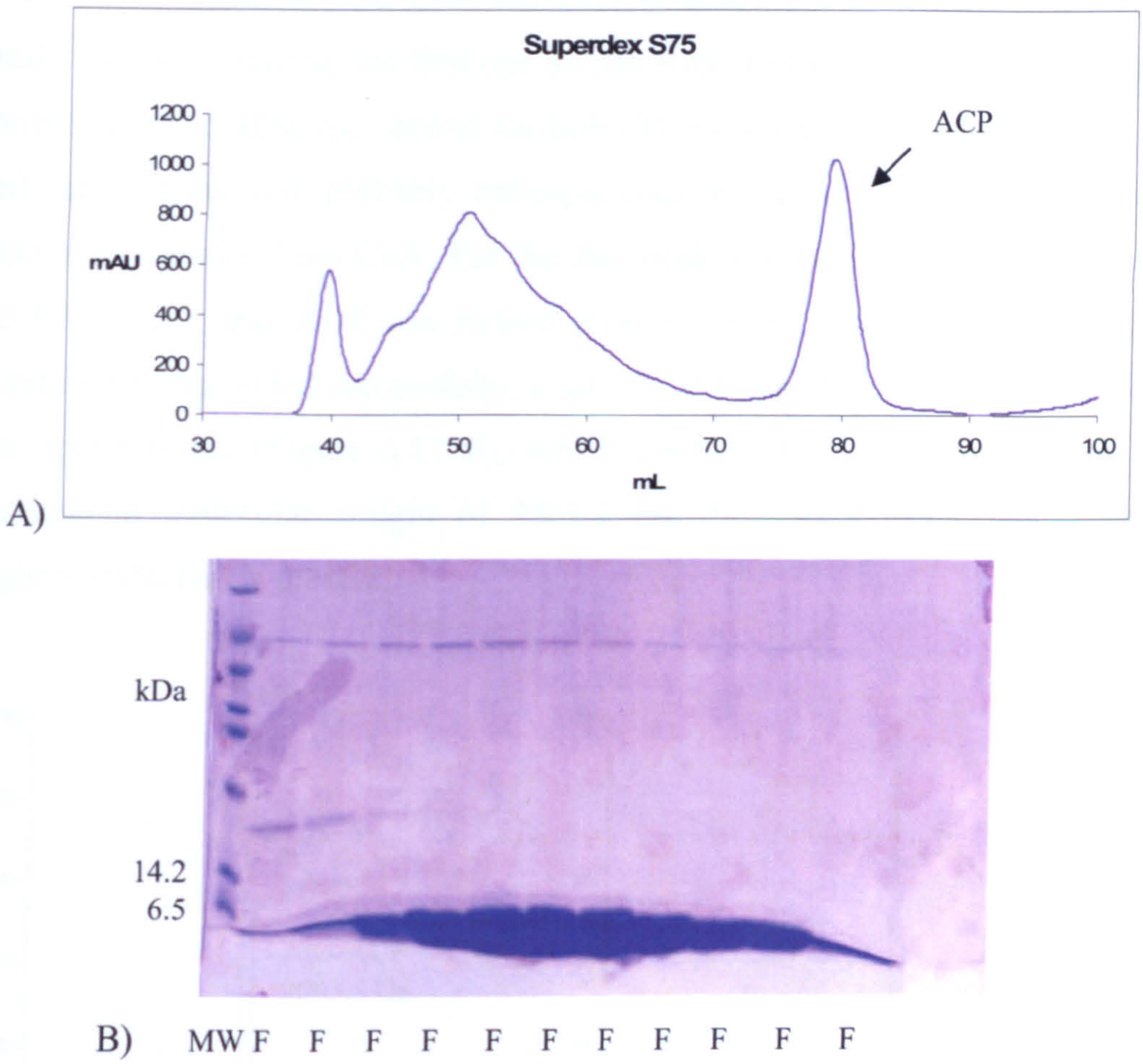


Figure 6.16 ACP gel filtration purification. Chromatogram (A) and SDS-PAGE gel (B) after ACP gel filtration. ACP is the last peak (A), and is very pure after this chromatography step (B). MW: molecular weights; F: fractions containing the pure apo-ACP eluted between 74 and 85 mL.

6.3.3 Activation of holo-ACP

The phosphopantetheinylation of apo-ACP can be achieved by incubating AcpS and ACP together in the presence of CoA and magnesium [61, 73] (chapter 3.6.9). This simple experiment is very useful to produce holo-ACP, which can be separated from the AcpS by gel filtration. Firstly a sample of purified ACP was tested by mass spectroscopy, to confirm that it was apo-ACP (Figure 6.17 A). As can be seen from the spectrum, the sample was quite homogeneous, with an

experimental molecular weight of 9094.8 Da (expected molecular weight for apo-ACP is 9101 Da). This apo-ACP had lost its methionine, as confirmed by N-terminal sequencing. After incubation with AcpS and gel filtration, holo-ACP can be harvested. As can be seen from the chromatogram and related gel (Figure 6.18), 3 peaks can be identified: the first one for the AcpS trimer of a molecular weight of approximately 45 kDa, the second for holo-ACP (molecular weight of about 9400 kDa), and the last one probably corresponding to 3',5'-ADP, the product of the pantetheine transfer from CoA. For the last peak, no bands were observed in the SDS-PAGE gel, and ADP can indeed absorb UV at 280 nm. To check if the activation had happened successfully, a sample of holo-ACP was checked again by mass spectroscopy (Figure 6.17 B), which confirmed the transfer. A peak with experimental molecular weight of 9461.2 Da (holo-ACP calculated molecular weight is 9456 Da) is present.

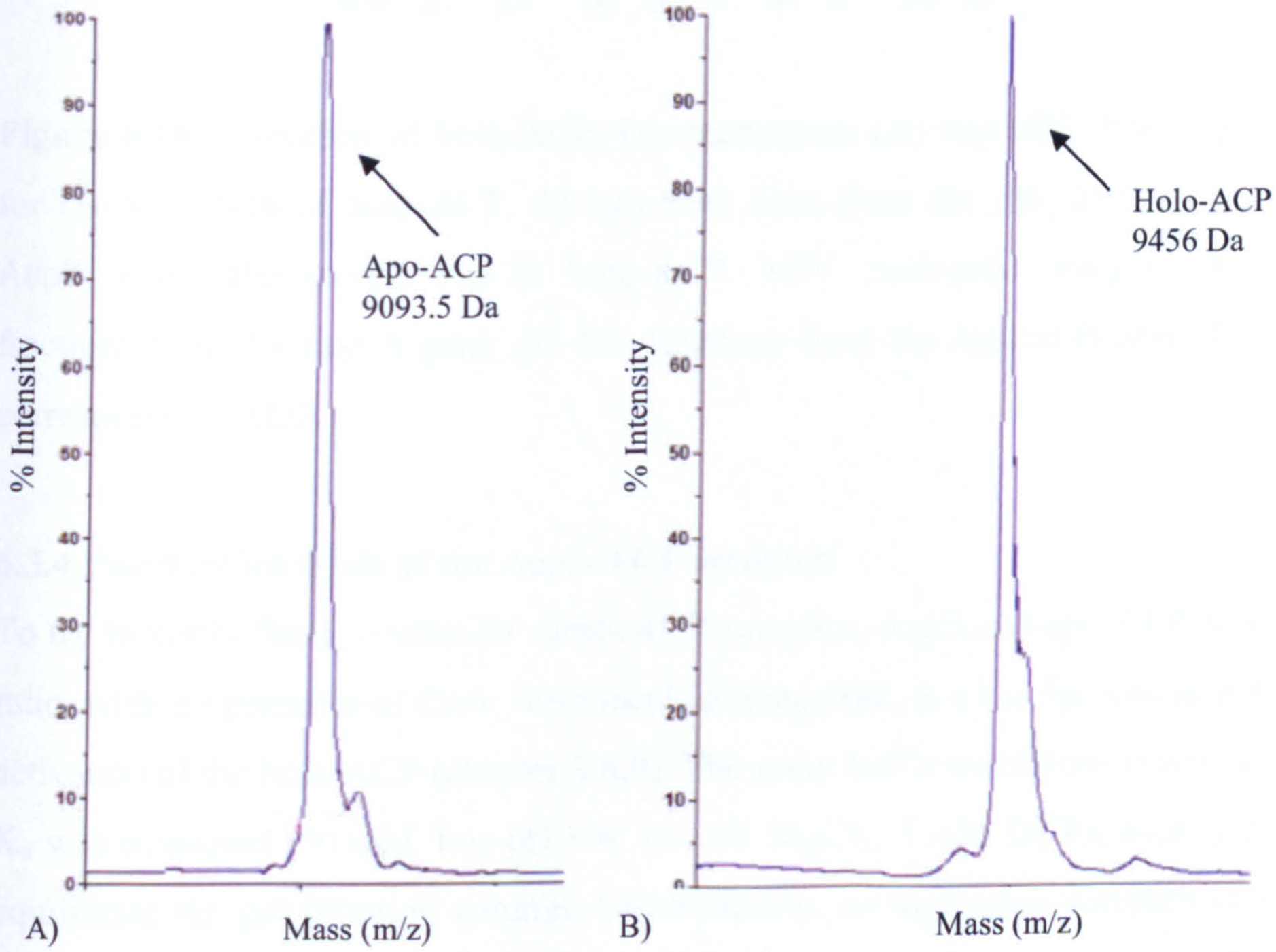


Figure 6.17 Mass spectra obtained for the apo-ACP (A) and holo-ACP (B). All errors are within ± 9 kDa.

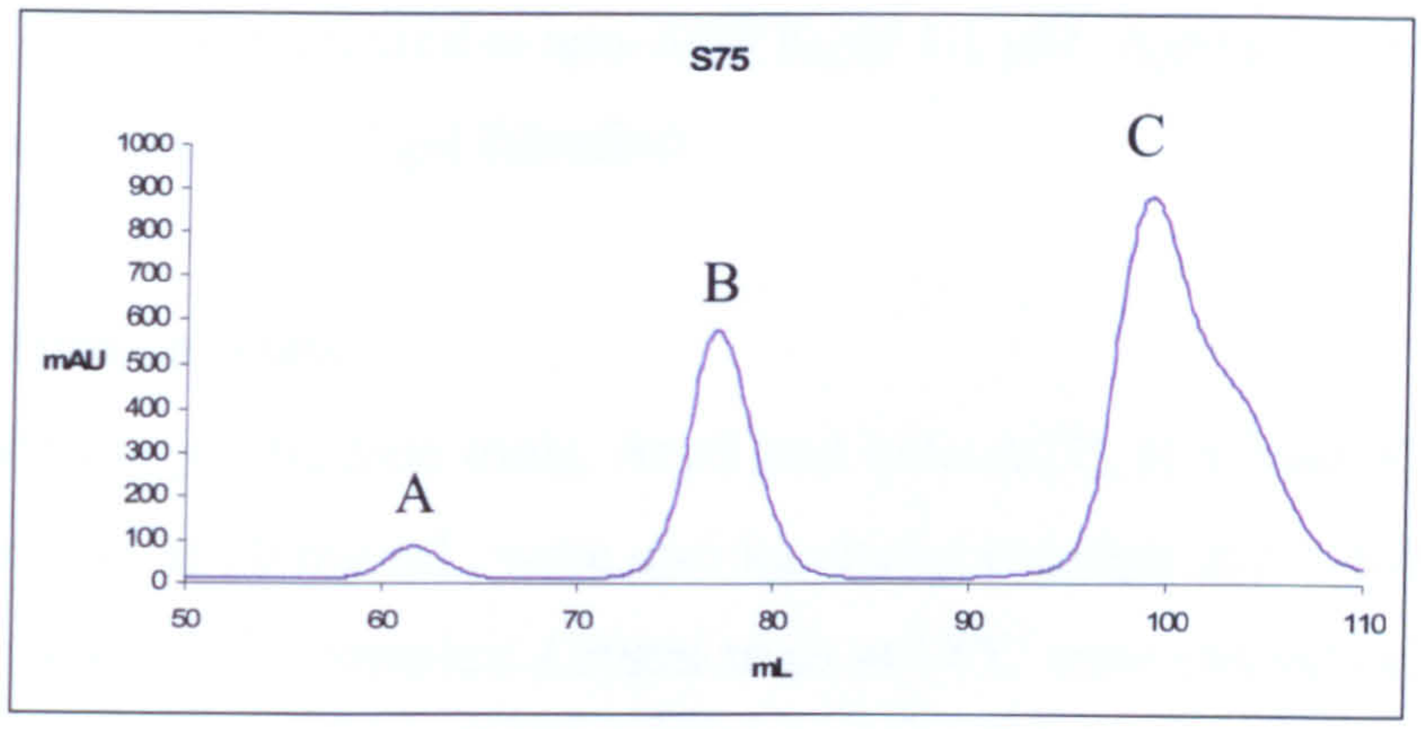


Figure 6.18 Activation of holo-ACP. Chromatogram (A) and SDS-PAGE gel (B) for the activation of holo-ACP. As can been seen from the gel, the first peak is AcpS, while the second one is holo-ACP. MW: molecular weights; A1-A2: fractions from the first A peak; B1-B7: fractions from the second B peak. Peak C corresponds to ADP.

6.3.4 Purification trials of the AcpS-ACP complex

To try to purify the *S. coelicolor* AcpS:ACP complex, AcpS and apo-ACP at a 1:1 ratio, with the presence of CoA, were incubated together, in a similar way as for the activation of the holo-ACP (chapter 3.6.9). The same buffer conditions under which K_d was measured (50 mM Tris pH 8.8, 10 mM $MgCl_2$, 5 mM DTT), were used to equilibrate the gel filtration column. Unfortunately, no activation complex can be observed after gel filtration. The resulted chromatogram and gels shows exactly the same peaks and bands that are obtained when apo-ACP is activated to holo-ACP (Figure 6.18). A peak is observed for both AcpS and ACP, but none for a putative complex (expected molecular weight of 72K, it should be a trimer of dimers). The same experiment was carried out incubating AcpS and the holo-ACP, since the

holo-ACP K_d is 2 μM, compared to apo-ACP K_d of 1.1 μM. Again, no complex was observed under conditions of gel filtration.

6.3.5 Crystallization trials

In parallel with the purification trials, AcpS and holo-ACP, at a ratio of 1:1, for a final concentration of 10 mg/mL, were also incubated together in a crystal drop, in order to co-crystallize the complex. Crystal trials at 18°C were carried out using the Clear Strategy Screen I and II, pH 6.5 (MDL), and the PEG/Ion Screen (Hampton). The presence of sodium cacodylate buffer with pH 6.5 was essential for the AcpS crystal growth, so it was chosen also for the co-crystallization trials hoping that it would help ACP to crystallize with the AcpS. The PEG/Ion Screen was chosen because the *B. subtilis* complex was crystallized using the same screen [86] and because again several conditions resulted in AcpS crystals. The best crystals, about 0.8×0.1×0.1 mm (Figure 6.19), were obtained in potassium bromide and sodium acetate (Table 6.2).

Salt	Buffer	PEG
0.2 M potassium bromide	0.1 M Sodium Cacodylate pH 6.5	25% PEG 2K MME
0.2 M potassium bromide	0.1 M Sodium Cacodylate pH 6.5	15% PEG 4K
0.2 M potassium bromide	0.1 M Sodium Cacodylate pH 6.5	10% PEG 8K + 1K
0.2 M potassium bromide	0.1 M Sodium Cacodylate pH 6.5	8% PEG 20K + 550 MME
0.3 M sodium acetate	0.1 M Sodium Cacodylate pH 6.5	15% PEG 4K
0.3 M sodium acetate	0.1 M Sodium Cacodylate pH 6.5	10% PEG 8K + 1K
0.3 M sodium acetate	0.1 M Sodium Cacodylate pH 6.5	8% PEG 20K + 550 MME

Table 6.2 Co-crystallization conditions for AcpS and holo-ACP.

Several crystals from both conditions were tested at the beamline 10.1 in Daresbury. Unfortunately, they turned out to be apo-AcpS crystals. As can be seen from Figure 6.19, these crystals are very similar to the apo-AcpS one. In order to check if both proteins were present, the worst diffracting crystals, with a difficult space group determination and data processing, were also resuspended in the crystallization buffer and the sample was run on an SDS-PAGE gel. As expected,

the gel showed the presence of only AcpS, suggesting that these crystals did not contain the ACP.

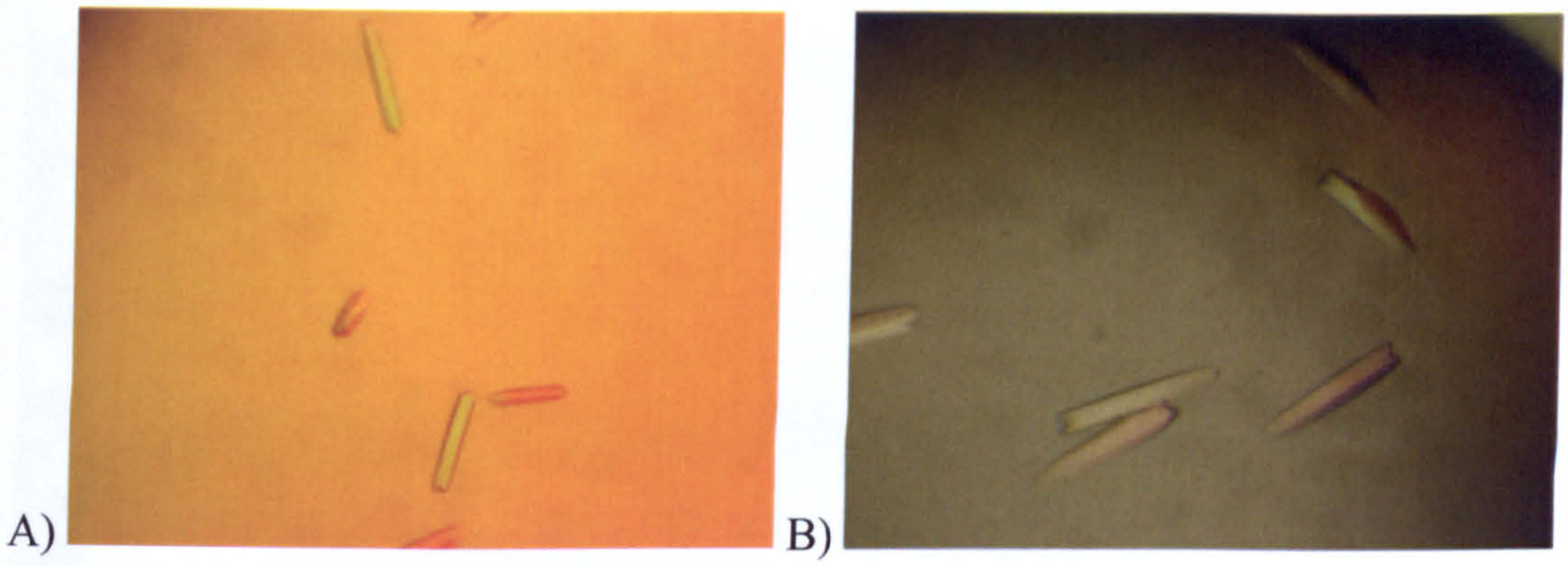


Figure 6.19 Crystals obtained after co-crystallization of AcpS and holo-ACP. A) Crystals obtained using 0.3 M sodium acetate, 0.1 M sodium cacodylate pH 6.5 and 15% PEG 4K as precipitant. B) Crystals obtained using 0.2 M potassium bromide, 0.1 M sodium cacodylate pH 6.5 and 15% PEG 4K as precipitant.

6.3.6 Thrombin cleavage of AcpS

The polyhistidine tag may affect the binding of ACP to AcpS, possibly even interacting with the negative helix α2 of ACP (the tag can have a positive charge in solution, depending on pH). It was observed that the polyhistidine tags come together at the top of the trimer axis in AcpS, and are close to helix α1, the proposed ACP binding site. It was also observed that in the case of *B. subtilis*, it was not possible to form a complex with native tagged or untagged AcpS. So, due to the expected similarity between the *S. coelicolor* and *B. subtilis* complexes, it was tried to remove the polyhistidine tag. The pET-15b-AcpS DNA contains a thrombin cleavage site, so attempts to cut the polyhistidine tag were carried out at two different temperature, 18ºC and 37ºC (chapter 3.6.6). As can be seen from the SDS-PAGE gel (Figure 6.20), a complete cut was obtained after 2 days for the 37ºC experiment and after 4 days for the 18ºC one. Unfortunately, the untagged protein precipitates. Only 0.3 mg from the initial 5 mg of protein could be recovered in the soluble fraction. The untagged AcpS was then purified from the his-AcpS (chapter 3.6.6) and it was incubated with apo- or holo-ACP, but again after gel filtration no complex was purified, and the corresponding chromatograms were similar to the one shown in Figure 6.18. This corresponds to what Parris observed [86]. It is

therefore possible that removing the histidine tag is not sufficient to isolate the AcpS:ACP complex.

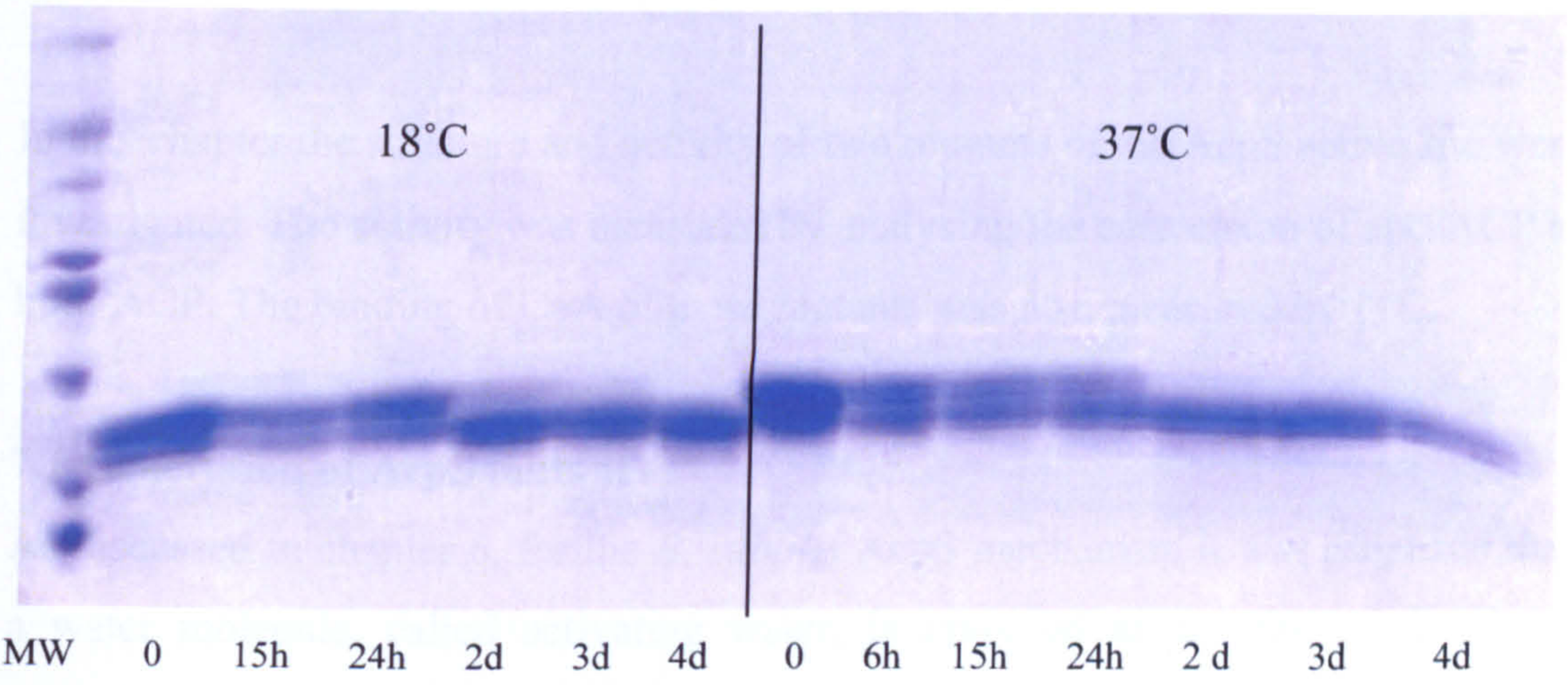


Figure 6.20 Thrombin cleavage of AcpS histidine tag. The SDS-PAGE gel was loaded with fractions collected after different amounts of time. Complete cleavage of the tag is obtained after 4 days at 18°C and 2 days at 37°C.

6.3.7 Conclusions

Several attempts to purify and crystallize *in vitro* the complex between AcpS and apo- or holo-ACP were carried out. Unfortunately, these experiments were not successful. The co-crystallization of the apo-AcpS and holo-ACP resulted in apo-AcpS like crystals, which contained only the AcpS. It was observed that AcpS helix α1, that interacts with ACP, moves in presence of CoA. This suggests that, in order to form a complex, holo-AcpS should be used. In addition, to facilitate the crystallization of a stable ternary complex of AcpS, CoA and ACP, the active serine in the ACP can be mutated to an alanine, or alternatively, a non-hydrolyzable CoA analogue can be used.

To avoid any aspecific interaction between the histidine tag of AcpS and helix α2 of ACP, attempts to cut the tag with thrombin cleavage were also carried out. Unfortunately, the untagged AcpS precipitates in solution, and a very small amount of soluble protein could be recovered. In order to increase the solubility of the untagged protein, different stabilizing buffers can be explored. Despite numerous attempts, Parris *et al*. managed to isolate the *B. subtilis* AcpS:ACP complex only when using a mutated AcpS that did not contain a histidine tag [86]. So, this approach can also be investigated for the *S. coelicolor* AcpS.

Chapter 7

Structure and activity of AcpS mutants

In this chapter the structure and activity of two mutants of the AcpS active site were investigated. The activity was measured by analysing the conversion of apo-ACP to holo-ACP. The binding of CoA of these mutants was also measured by ITC.

7.1 Generation of AcpS mutants

As discussed in chapter 6, for the *B. subtilis* AcpS mechanism it was proposed that a water molecule, called activating water, is involved at the beginning of the phosphopantetheine transfer, deprotonating the conserved ACP serine.

Two residues, His110 and Asp111, were identified in the *S. coelicolor* AcpS to be the possible catalytic residues responsible for direct deprotonation of the ACP serine, or base catalysts in the deprotonation of a water molecule. These residues are placed on either side of the β-phosphate of CoA, and after inspection of the *B. subtilis* AcpS:ACP complex and *S. coelicolor* model, it is possible to suggest that they could both reach the ACP serine and deprotonate it.

At the pH where the phosphopantetheinylation occurs, the carboxyl group of the aspartate side chain is deprotonated, and its oxygens highly electronegative can remove a proton from the hydroxyl group of the serine. Asp111 could act in a similar way to the *B. subtilis* ACP Asp35 [86], but allowing less specific binding to promote a broader specificity for ACP recognition, without the need of an additional base contributed by ACP which would require precise positioning. In addition, Asp111 is the residue that has the greatest change in position on CoA binding. Whilst the aspartate residue is not completely conserved among all the other AcpS, the histidine 110 is highly conserved (Figure 7.1). The basic nitrogen of the imidazole ring of histidine is usually involved in a quick proton shuttle, and it can remove a proton from serine, threonine or cysteine residues activating them as nucleophiles.

To explore their role during catalysis, these two residues were mutated to alanines by site-directed mutagenesis, and the correspondent proteins were expressed,

purified and crystallized. Finally, their structures were solved by molecular replacement with the wild type. ITC was also used to calculate K_d of CoA binding.

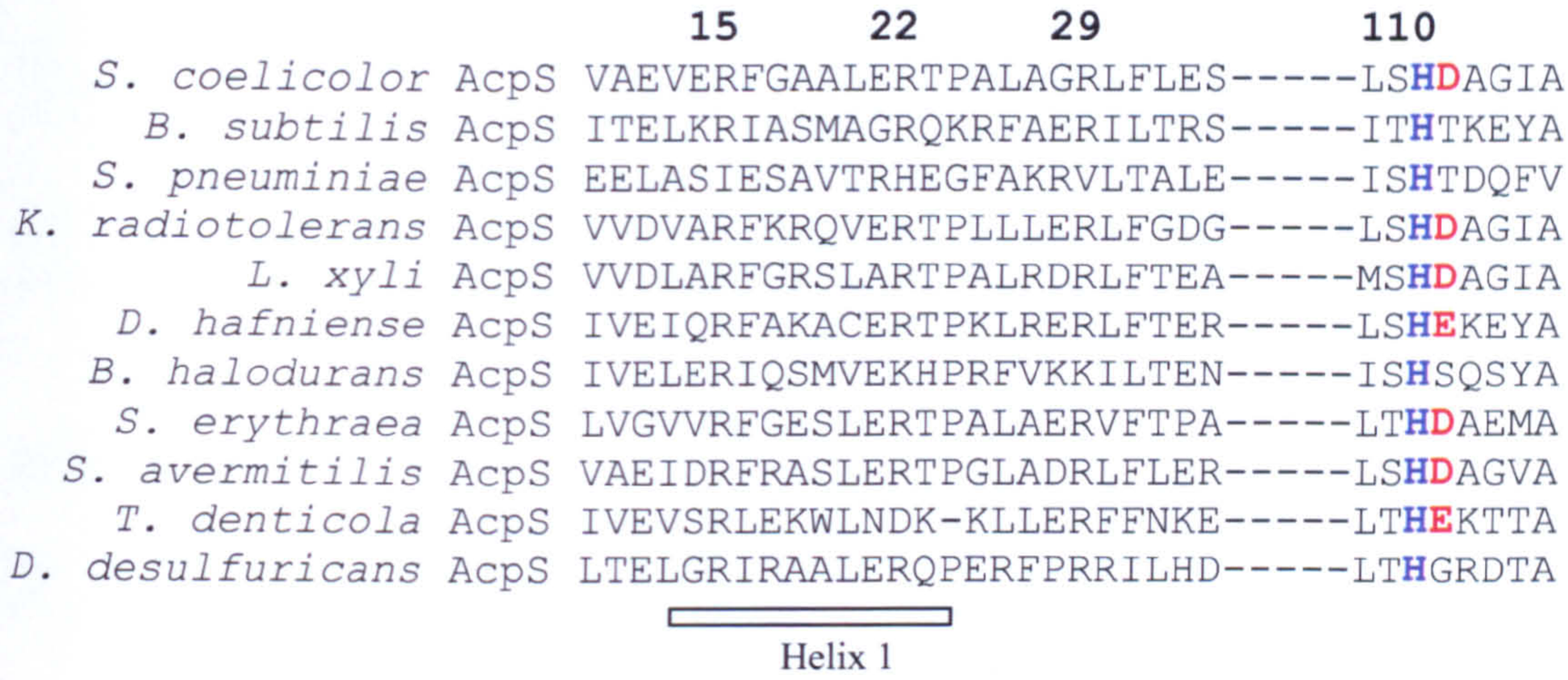


Figure 7.1 Alignment of multiple polyketide synthesis and fatty acid synthesis Acyl carrier protein Synthase sequences.

7.2 Primers design and cloning

The Stratagene QuickChange II Site-Directed Mutagenesis kit was used for the generation of the plasmids containing the single mutation. This kit was chosen for its simplicity of use and for the relatively short time it requires. The only change to the protocol provided with the kit (chapter 3.4.4) was the addition of DMSO. Due to the high GC content (more than 75%), GC hairpins in the DNA could be formed, and DMSO was essential to destroy this base pairing and allow PCR reaction to occur. *E. coli* BL21 (DE3) cells were transformed with the new plasmids, and after miniprep the mutated vector was sent for sequencing, to confirm the presence of the single mutation. As can be seen from Figure 7.2, the site-directed mutagenesis was successful and the Asp111 or the His110 were substituted with alanine residues, and no other mutations occurred into the gene or the histidine tag.

A) D111A mutant

```
wt     METGSSHHHHHHSSGLVPRGSHMETSIIGVGIDVAEVERFGAALERTPALAGRLFLESEL
D111A  METGSSHHHHHHSSGLVPRGSHMETSIIGVGIDVAEVERFGAALERTPALAGRLFLESEL
       ************************************************************

wt     LLPGGERRGVASLAARFAAKEALAKALGAPAGLLWTDAEVWVEAGGRPRLRVTGTVAARA
D111A  LLPGGERRGVASLAARFAAKEALAKALGAPAGLLWTDAEVWVEAGGRPRLRVTGTVAARA
       ************************************************************

                        111
wt         AELGVASWHVSLSHDAGIASAVVIAEGSTOP
D111A      AELGVASWHVSLSHAAGIASAVVIAEGSTOP
           ********** **********
```

B) H110A mutant

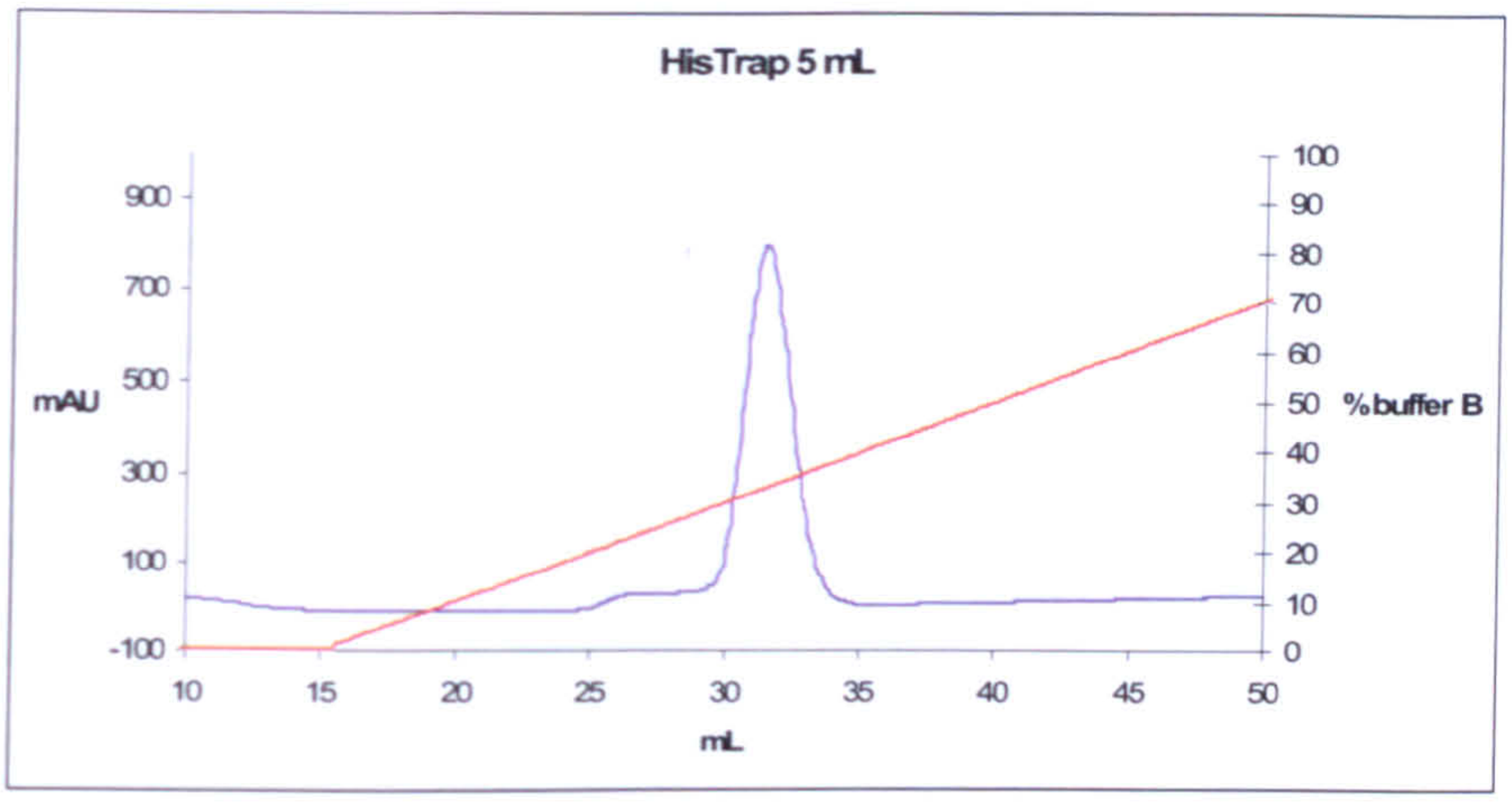
```
wt     METGSSHHHHHHSSGLVPRGSHMETSIIGVGIDVAEVERFGAALERTPALAGRLFLESEL
H110A  METGSSHHHHHHSSGLVPRGSHMETSIIGVGIDVAEVERFGAALERTPALAGRLFLESEL
       ************************************************************

wt     LLPGGERRGVASLAARFAAKEALAKALGAPAGLLWTDAEVWVEAGGRPRLRVTGTVAARA
H110A  LLPGGERRGVASLAARFAAKEALAKALGAPAGLLWTDAEVWVEAGGRPRLRVTGTVAARA
       ************************************************************

                      110
wt         AELGVASWHVSLSHDAGIASAVVIAEGSTOP
H110A      AELGVASWHVSLSADAGIASAVVIAEGSTOP
           ********** **********
```

Figure 7.2 Alignments between the wild type AcpS and the D111A (A) or H110A (B) mutant. Nucleotide sequences after sequencing were translated with Translate (available at the expasy website [149]), then aligned with ClustalW [148] ("*" means identical residue).

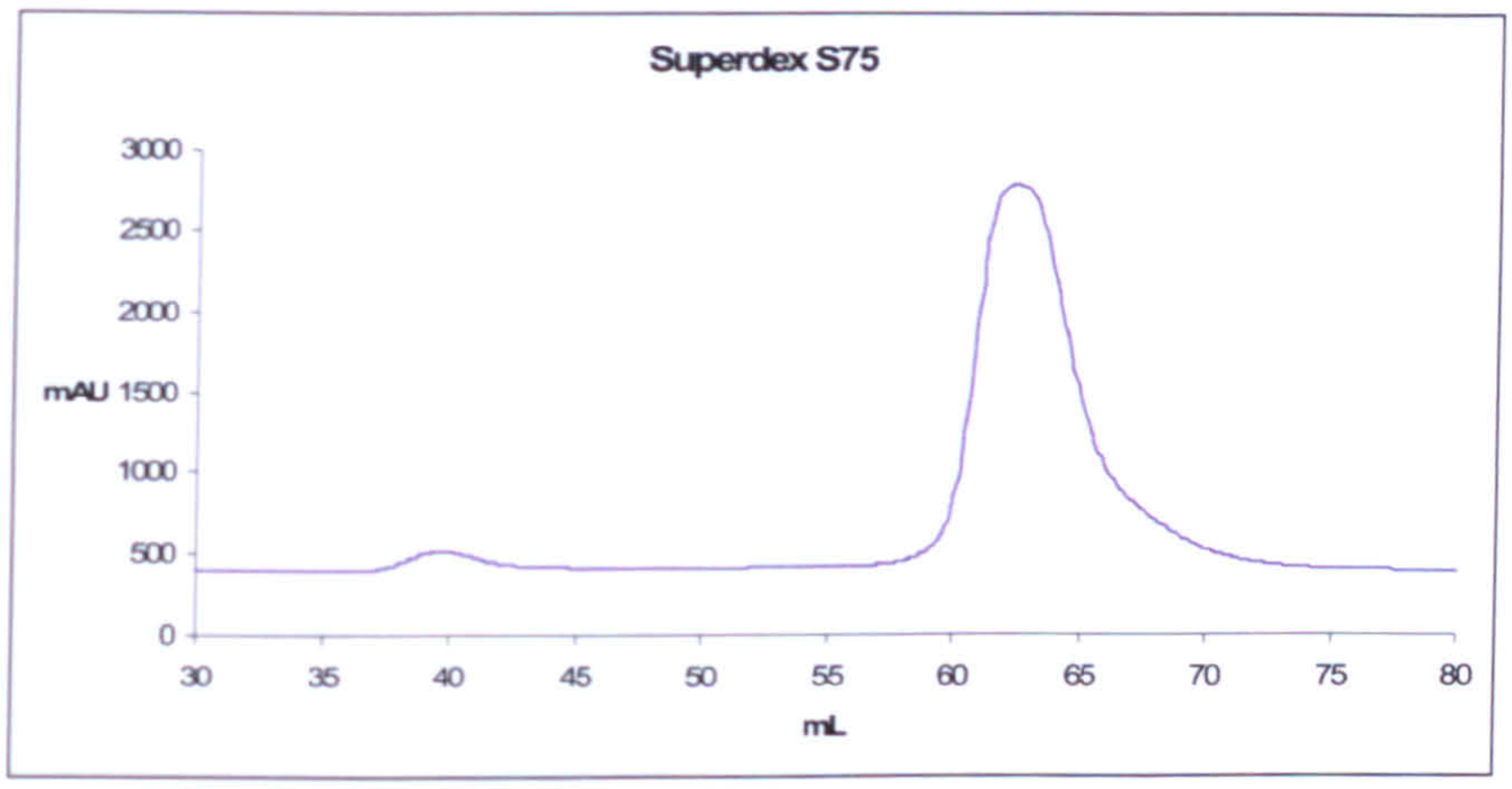
7.3 Expression and purification of H110A and D111A mutants

Both mutants were expressed and purified using the same protocol as the wild type AcpS: a nickel affinity chromatography (chapter 3.6.3), an optional gel filtration column (chapter 3.6.5) and desalting (chapter 3.6.8). Apart from the fact that both mutants are eluted at a slightly higher imidazole concentration, about 300 mM from the nickel column, no further differences in comparison with the wild type were observed (Figure 7.3). About 10 mg of pure protein were obtained from 1 L of bacterial culture.

A)



B)



C)

Figure 7.3 Purification of AcpS mutants. Examples of affinity (A), gel filtration chromatography (B) and SDS-PAGE gel (C) for H110A and D111A AcpS mutants. UV absorbance is shown in blue, the gradient of the elution buffer in red. Mutants are eluted at a slightly higher concentration of imidazole, and no further differences are observed during their purification in comparison to the wild type.

7.4 Investigation of AcpS mutant activities on ACP by mass spectroscopy

Each mutant was incubated with the polyketide *S. coelicolor* ACP and a standard phosphopantetheine transfer reaction was set up *in vitro* (chapter 3.6.9). Each reaction was carried out using the same amount of protein and time (1h). The resulting ACP was then purified by gel filtration (Figure 7.4) and mass spectroscopy was used as initial investigation to determine whether these mutants were still capable of phosphopantetheinylation (apo- and holo-ACP differs by 300 Da, difference that is too small to be seen during size exclusion separation).

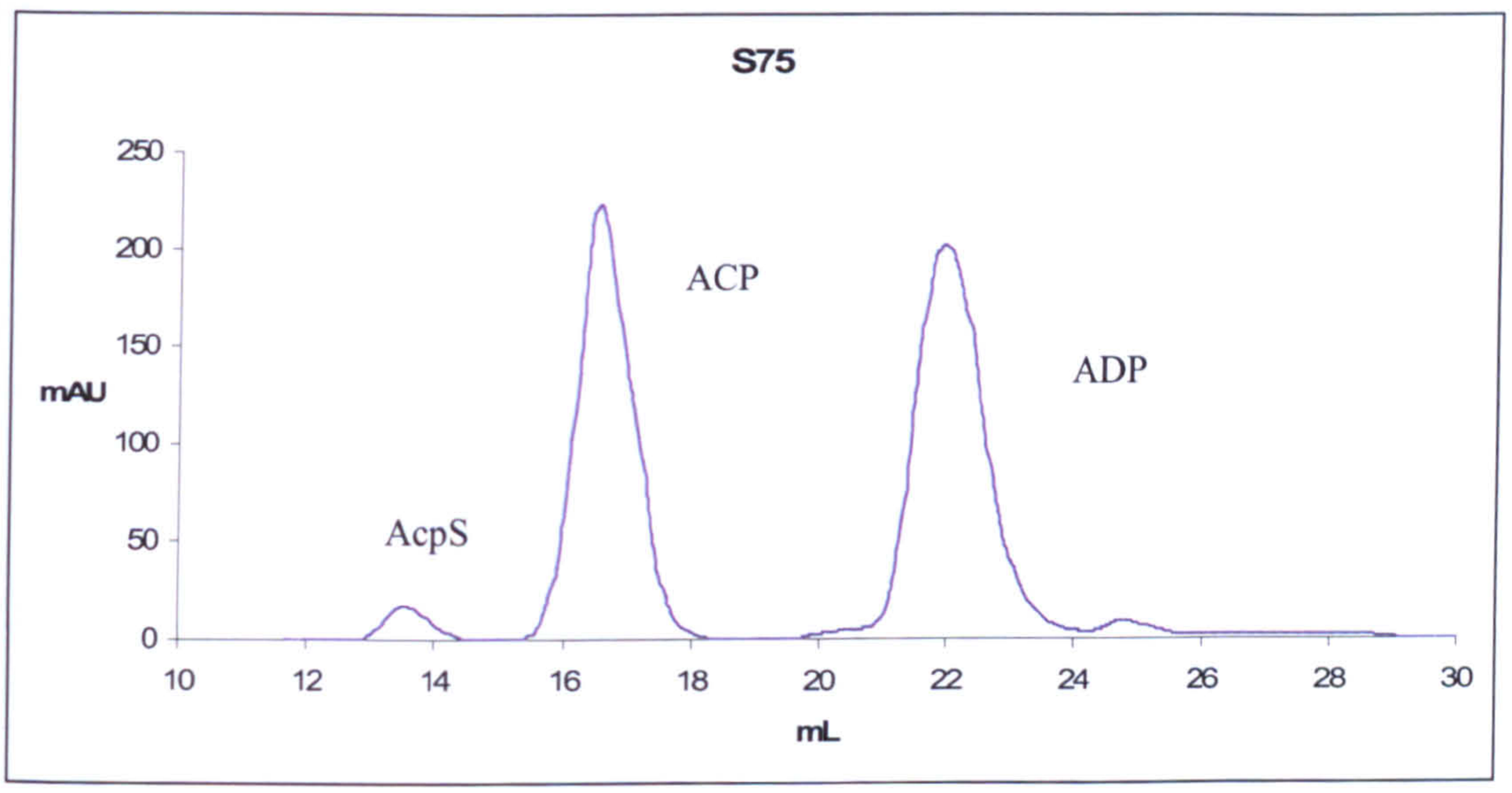


Figure 7.4 Purification of holo-ACP by gel filtration. AcpS and ACP were incubated together at 1:100 ratio with the presence of CoA.

A control reaction with wild type AcpS was also carried out. If activity was preserved, ACP should be present in its holo form. As can be seen from the mass spectra, the incubation with wild type AcpS led to the formation of holo-ACP. On the contrary, with the aspartate mutant, by comparing the intensity of the peaks, about 55% of holo-ACP seems to be present in solution, with 45% reduction in activity (Figure 7.5). The histidine mutation seemed to cause a greater effect on catalysis. The activity was reduced by approximately the 85%. So, it is evident that both mutants cause a reduction of the phosphopantetheine transfer, but with this analysis it is not possible to determine if this loss in activity is due to a catalytic role for these residues, or because CoA was not bound, suggesting a structural role

for them. To investigate the role of the His110 and Asp111 further, their crystal structures were solved in the presence of CoA for comparison to the wild type AcpS. In addition, ITC was also essential to determine the cofactor binding constant.

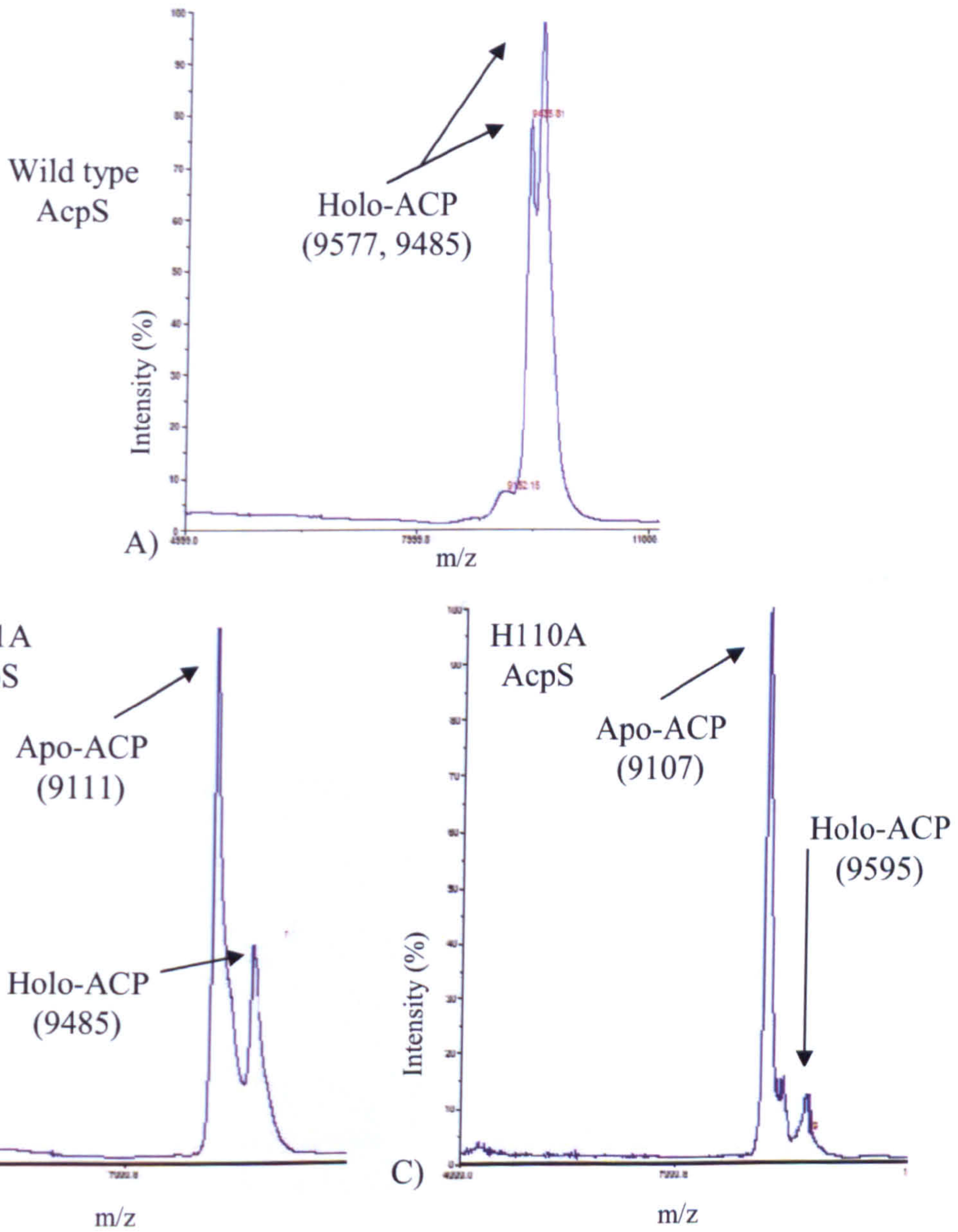


Figure 7.5 Activation of ACP by mutants. ACP Mass spectra measured on ACP samples obtained after phosphopantetheinylation reaction with wild type AcpS, D111A and H110A mutants. All errors are within ± 9 kDa, and several sodium ions are bound to the ACPs.

7.5 Determination of AcpS D111A structure

7.5.1 AcpS D111A crystallization

Solutions from the MDL Clear Strategy Screen I and II, at pH 6.5, and the Hampton PEG/Ion Screen, with a protein concentration of 10 mg/mL and in the presence of 5 mM CoA, were screened at 18˚C for D111A crystals (chapter 3.8.1). These screens were chosen because they gave the best results for the wild type AcpS. The best crystals, measuring about 0.7×0.6×0.3 mm (Figure 7.6) were obtained in the conditions shown in Table 7.1:

Salt	Buffer	PEG
0.15 M potassium thiocyanate	0.1 M sodium cacodylate pH 6.5	15% PEG 4K
0.2 M potassium bromide	0.1 M sodium cacodylate pH 6.5	15% PEG 4K
0.2 M potassium thiocyanate	0.1 M sodium cacodylate pH 6.5	8% PEG 20K + 8% PEG 550 MME
0.2 M tri-lithium citrate	none	20% PEG 3350
0.2 M lithium decahylate	none	20% PEG 3350

Table 7.1 D111AcpS crystallization conditions.

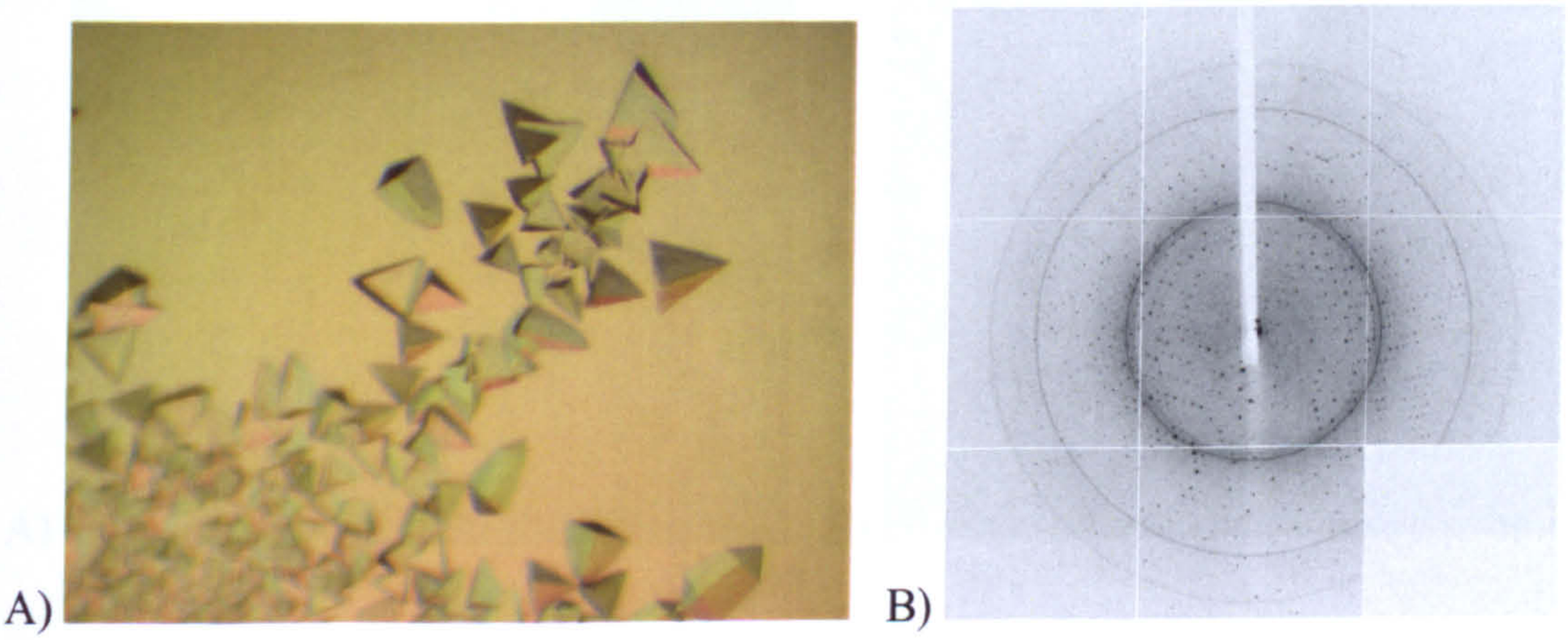


Figure 7.6 Example of D111A AcpS crystal (A) and its diffraction image (B). One of the CCDs of the detector was not working during data collection at IO2 in Diamond. This area was excluded during integration. Crystals were obtained in 0.2 M potassium thiocyanate, 0.1 M sodium cacodylate pH 6.5, 8% PEG 20K + 8% PEG 550 MME.

7.5.2 Data collection and processing

Several datasets were collected from each condition at beamline IO2 in Diamond. The best crystal, which diffracted to a resolution of approximately 1.4 Å, was obtained in 0.2 M potassium thiocyanate, 0.1 M sodium cacodylate pH 6.5, 8% PEG 20K + 8% PEG 550 MME. Glycerol was added to the same reservoir for a final concentration of 15% v/v, and this solution was used as cryoprotectant. The crystal was a bit icy as can be seen in picture 7.6 (presence of the characteristic ice ring at about 3.2 Å), but this did not affect data collection and processing. The crystal belonged to the $P2_13$ space group with a=b=c=72.8 Å.

7.5.3 Molecular replacement and refinement

One molecule per asymmetric unit was found. The structure was solved by molecular replacement using Phaser [136, 137] and the holo-AcpS as search model. Inspection of the difference map clearly showed the presence of the mutation (Figure 7.7). The aspartate was then mutated to an alanine in Coot [142] and the refinement was carried out with Refmac5 [133] (chapter 3.8.6).

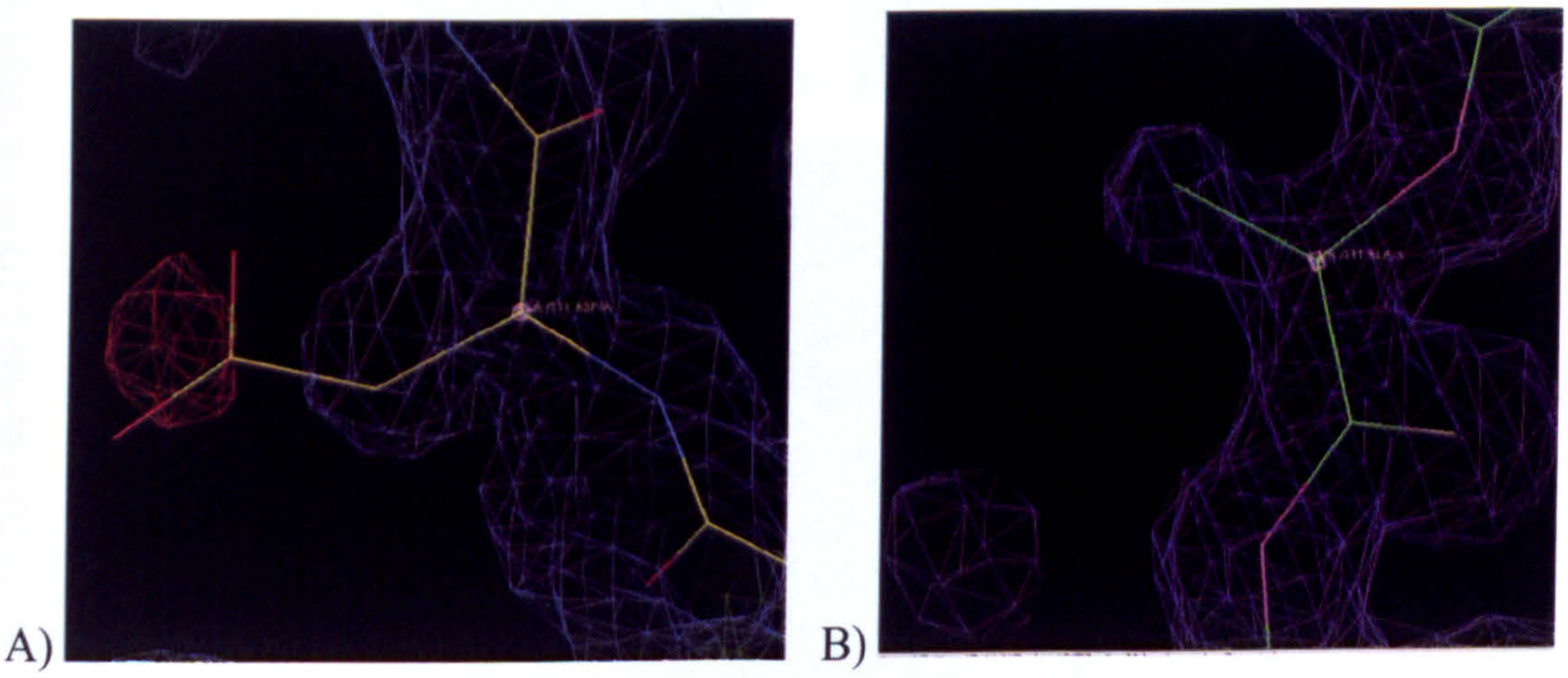


Figure 7.7 Electron density corresponding to the D111A mutation. Examination of the difference map (DELFWT, DELPHWT) related to the D111A mutant (A) clearly show the mutation. B) Density map corresponding to the alanine residue in the final model. The density map is contoured at 1σ, while the difference map at 3σ.

The model contains 123 residues, 2 magnesiums with 0.5 occupancy each, 2 molecule of CoA with 0.5 occupancy each, 1 sodium and 208 water molecules. Inspection of the electron density at the CoA binding site after refining the model with a full occupancy cofactor in a orientation similar to that one of the holo-AcpS (Figure 7.8 A), and the presence of two magnesium ions, suggested that CoA was bound in two alternative conformations, that begin to diverge from the α-phosphates. Whilst the electron density for the adenine rings of both CoAs, including the α- and β-phosphates, and their related magnesium ions was very clear, the density for the phosphopantetheine arms was not well defined and the model building for that part of the CoA was difficult (Figure 7.8).

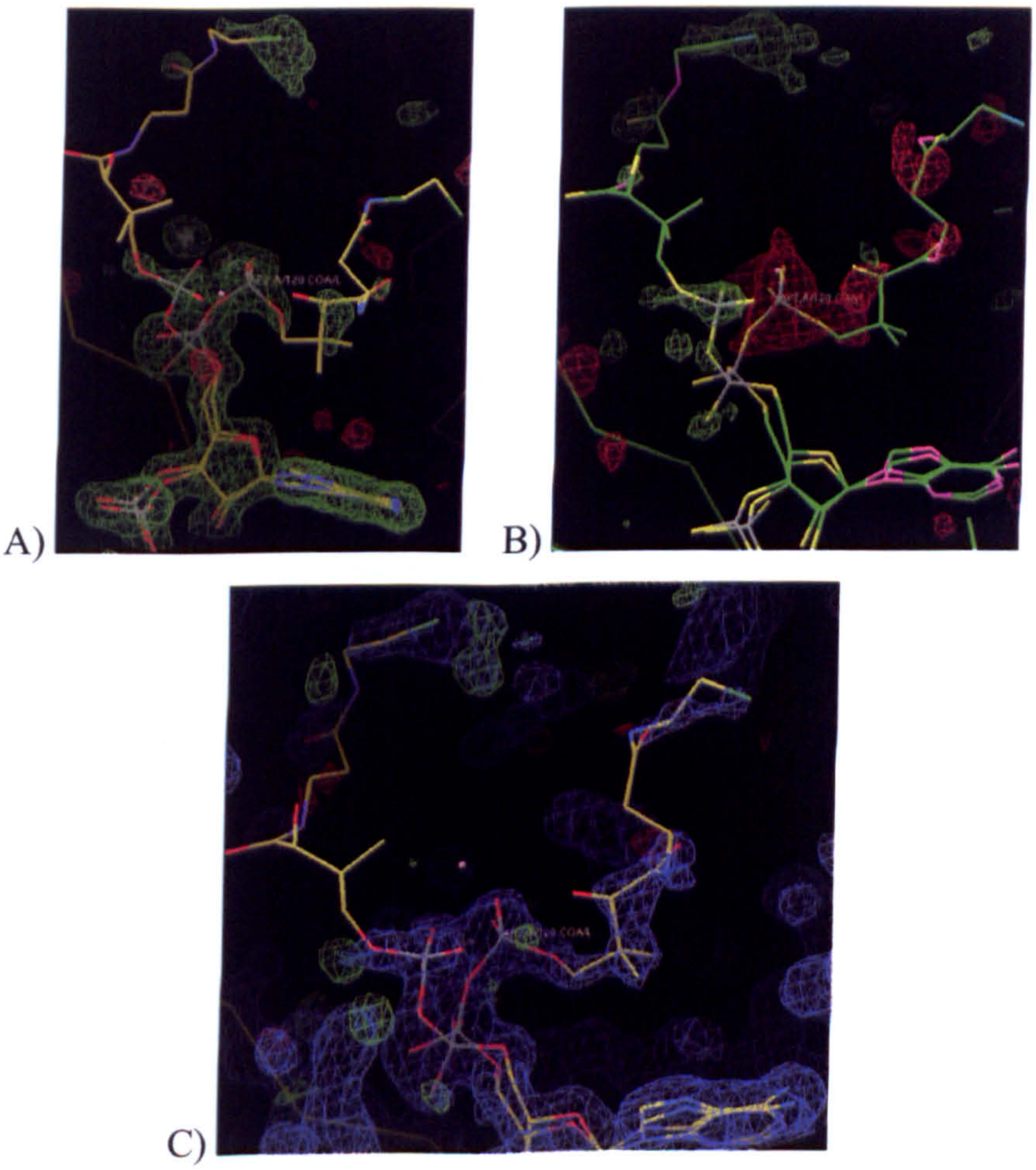


Figure 7.8 Possible alternative conformations of CoA bound to D111A AcpS. Difference maps (DELFWT, DELPHWT) related to the model refined without the cofactor (A), and related to the model refined with one molecule of cofactor with full occupancy (B). C) Electron density map (FWT, PHWT) and difference map (DELFWT, DELPHWT) of the final orientations of CoAs. The electron density map is contoured at 1.2σ, the difference map at 3σ. In figure 7.8 A and B the final model is shown as bonds.

The final model represents the clearly defined alternative positions of the adenine ring and phosphate for both CoAs, and for each model, the position for the phosphopantetheine chain was chosen which resulted in the least negative electron density. However there is evidence for other possible orientations of the phosphopantetheine chain portion, in contrast with the wild type structure where the density for the adenine base and the phosphopantetheine chain was all well defined. The presence of two binding sites for the cofactor was also proved by ITC experiment, as described in chapter 7.9. One molecule of CoA resulted to be bound in a similar way as in the holo-AcpS, while the second one to adopt a completely new conformation and is located in a different hydrophobic pocket. The refinement led to a structure model with R-factor an R_{free} values of 0.19 and 0.21 respectively. Statistics for data processing and refinement are shown in Table 7.2.

	D111A AcpS
Wavelength (Å)	0.953
Resolution (Å)	50-1.4 (1.45-1.4)
Space group	P 2_13 a=b=c=72.8 Å
Completeness (%)	99.6 (96.4)
N unique reflections	25707
Redundancy	7.9 (2.9)
I/σI	34.7 (1.5)
R merge	0.05 (0.51)
Mosaicity (°)	0.21
R_{free}	0.21
R-factor	0.19
RMS bond (Å)	0.08
RMS angle (°)	1.271

Table 7.2 Data processing and refinement statistics for D111A AcpS mutant. Data corresponding to the highest resolution shell are in brackets.

7.5.4 Geometry and B-factor analysis

The geometry and the conformation were checked using MolProbity [145, 146] . The Ramachandran plot is presented in Figure 7.9. All 123 residues are in the most favoured region. Only His110 was recognized to have a slightly unusual rotamer (0.4%). The imidazole ring is coordinated to the 3'-phosphates of the two alternative conformations of the cofactor. The O8 oxygens of these phosphates are 0.7 Å apart. So, the electron density related to the histidine is an average of two conformers, which explains the reason of the poor geometry.

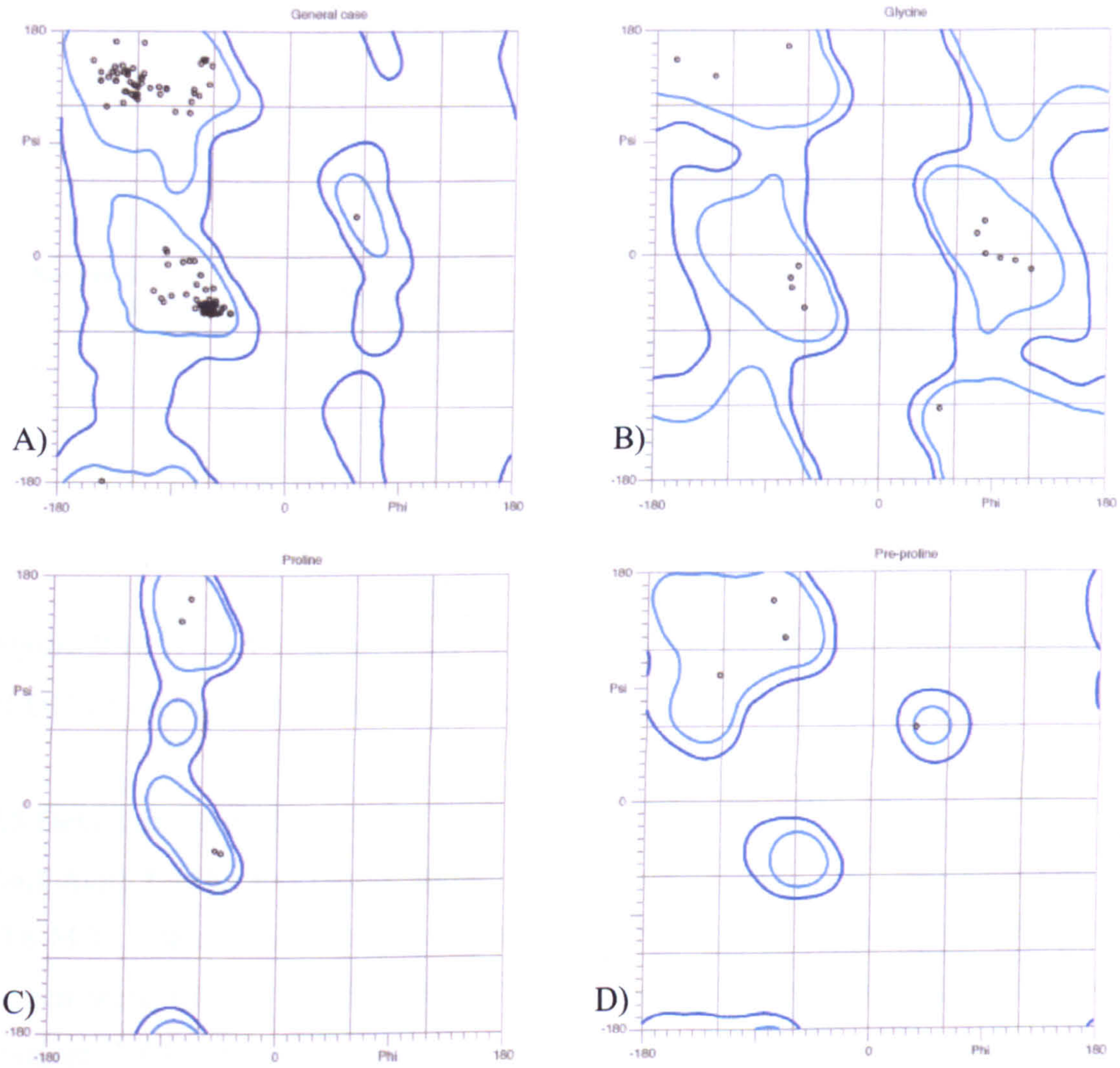


Figure 7.9 Ramachandran plot related to D111A mutant made by MolProbity. A) general case; B) glycine; C) proline; D) pre-proline.

An analysis of the average B-factor of the main chains was also performed with Baverage [143]. The average B-factor is 16 $Å^2$, while the average temperature factor for the mobile areas is higher: 21 $Å^2$ (MH1), 30 $Å^2$ (ML1) 17 $Å^2$ (ML2) and 28 $Å^2$ (ML3). The loop containing the Asp111, shifted from the apo-position, has also higher average B-factor: 19 $Å^2$. This B-factor analysis and a comparison with the wild type AcpS are shown in Figure 7.10. The B-factor of this mutant is lower than the wild type one, but its pattern is very similar, with higher temperature factor values for mobile loops.

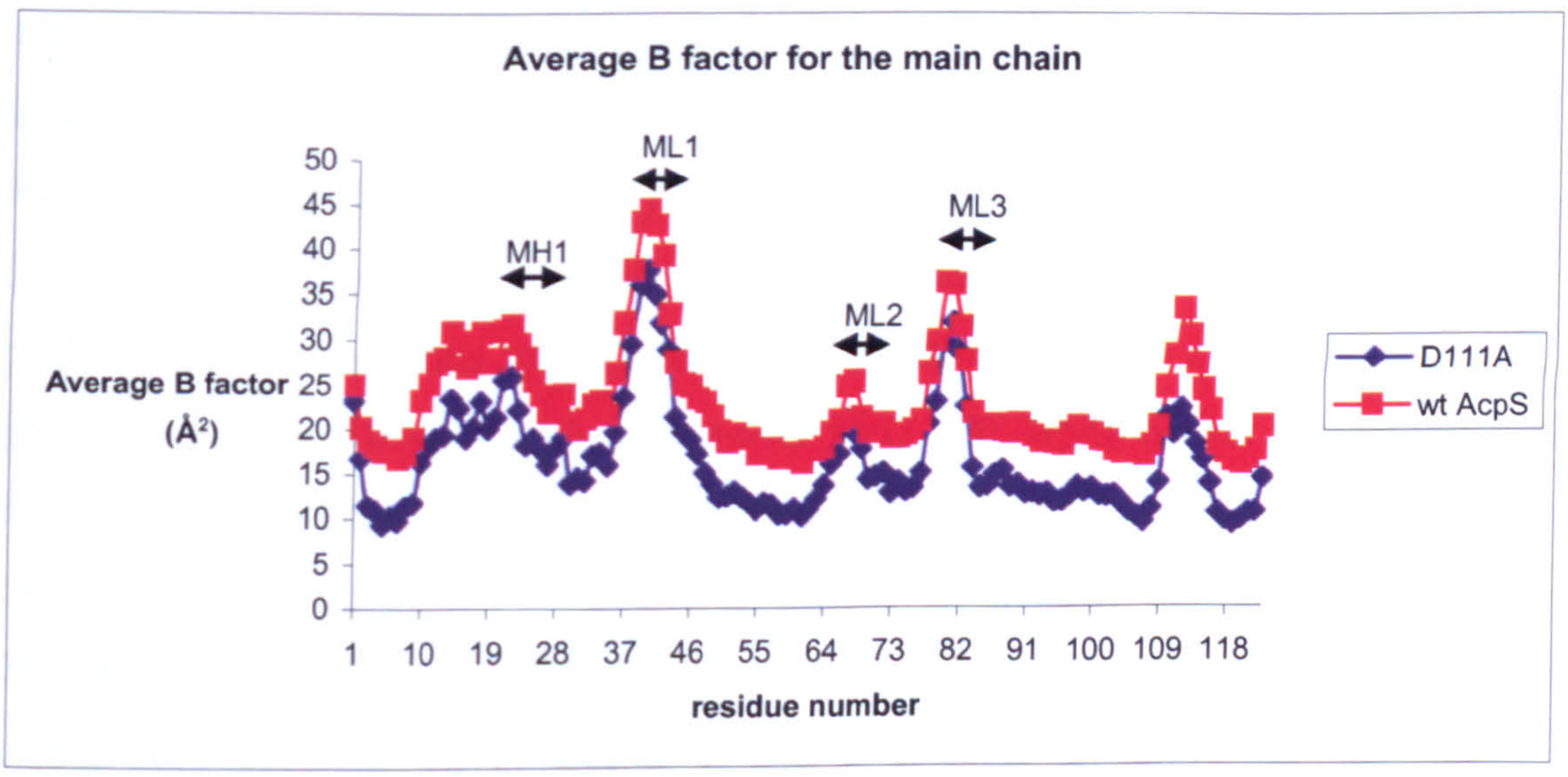


Figure 7.10 D111A B-factor analysis. Average B-factor values for the main chain of D111A mutant and its comparison with the wild type AcpS.

7.6 Determination of AcpS H110A structure

7.6.1 AcpS H110AA crystallization

The MDL Clear Strategy Screen I and II, at pH 6.5, and the Hampton PEG/Ion Screen were used to obtained H1110A crystals in the presence of 5 mM CoA (chapter 3.8.1). The best conditions, which gave the more regular and bigger crystals measuring approximately 0.8×0.6×0.3 mm at 18°C (Figure 7.11), are shown in Table 7.3.

Salt	Buffer	PEG
0.3 M sodium acetate	0.1 M sodium cacodylate pH 6.5	25% PEG 2K MME
0.15 M potassium thiocyanate	0.1 M sodium cacodylate pH 6.5	18% PEG5K MME
5 mM Nickel chloride	0.1 M sodium cacodylate pH 6.5	20% PEG 4K
0.2 M magnesium chloride hexahydrate	none	20% PEG 3350
0.2 M tri-lithium citrate tetrahydrate	none	20% PEG 3350

Table 7.3 Crystallization conditions for the H110A AcpS mutant.

7.6.2 Data collection and processing

Crystals from each condition were tested at beamline BM10.1 at the Synchrotron Radiation Source (SRS), Daresbury, and IO2 and IO3 at the Diamond Light source (DLS). It was noted that, the bigger the crystal was, the worse its diffraction was, probably due to problems while freezing. In addition, the best cryosolutions were proved to contain 10%-15% v/v glycerol or PEG 400, while higher concentrations destroyed the crystals. The best dataset was collected at station 10.1 in Daresbury, from a crystal obtained in 0.3 M sodium acetate, 0.1 M sodium cacodylate pH 6.5, 25% PEG 2K MME with the presence of 10% PEG 400 as cryoprotectant. The crystal belonged to the $P2_13$ space group with a=b=c=72.7 Å.

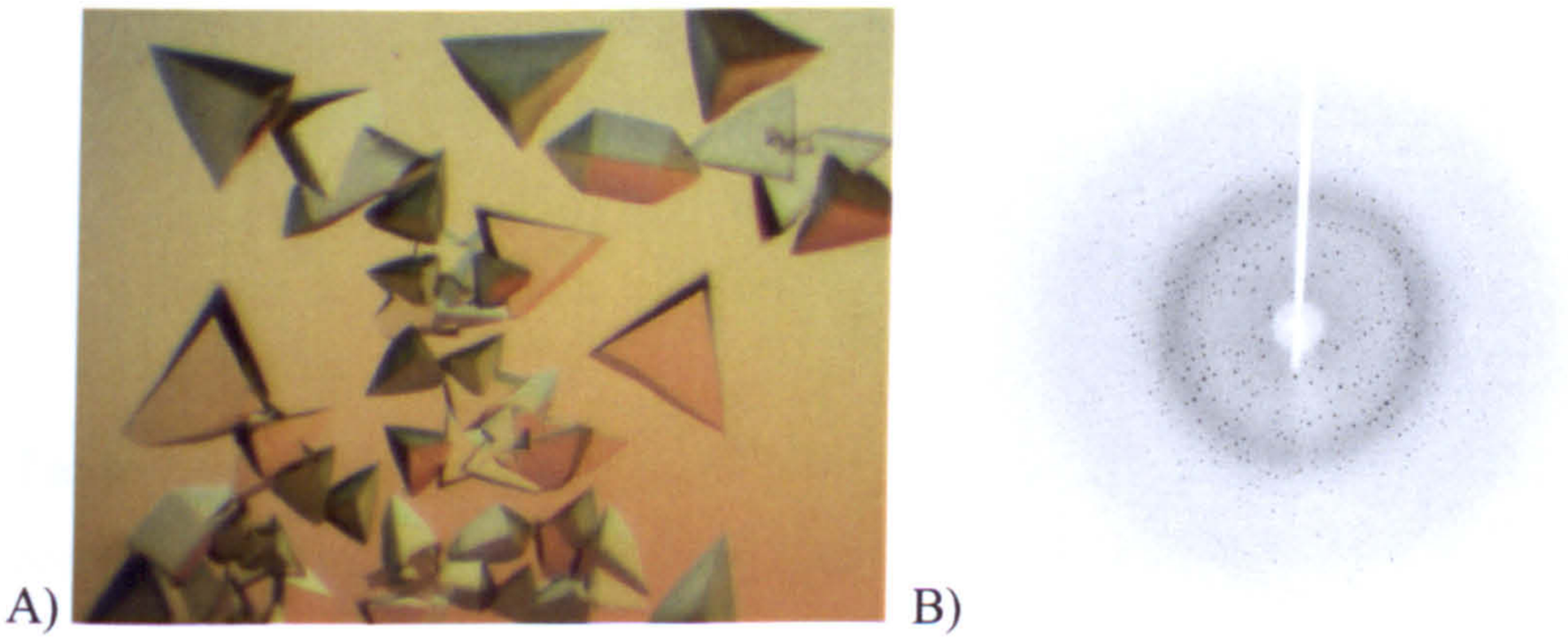


Figure 7.11 Example of H110A AcpS crystal (A) and its diffraction image (B). Crystals were obtained in 0.3 M sodium acetate, 0.1 M sodium cacodylate pH 6.5, 25% PEG 2K MME.

7.6.3 Molecular replacement and refinement

The structure was solved by molecular replacement using Phaser [136, 137] and the wild type holo-AcpS as search model. Examination of the difference map showed the presence of the mutation, even though this identification was not immediate, due to the fact that a water molecule took the position of the histidine side chain. One molecule per asymmetric unit was found. The model contains 123 residues, 1 magnesium, 1 molecule of CoA and 237 water molecules. The refinement led to a structure model with R-factor an R_{free} values of *0.18 and 0.20 respectively.* Data processing and refinement statistics are shown in Table 7.4.

	H110A AcpS
Wavelength (Å)	1.117
Resolution (Å)	50-1.35 (1.40-1.35)
Space group	P $2_1$3 a=b=c=72.7 Å
N unique reflection	28420
Completeness (%)	99.3 (99.3)
Redundancy	9.3 (2.1)
I/σI	32 (2.5)
R merge	0.07 (0.22)
Mosaicity (°)	0.41
R_{free}	0.20
R-factor	0.18
RMS bond (Å)	0.007
RMS angle (°)	1.229

Table 7.4 Data processing and refinement statistics for D111A AcpS mutant. Data corresponding to the highest resolution shell are in brackets.

7.6.4 Geometry and B-factor analysis

The geometry and the conformation were checked using the tool of programs MolProbity [145, 146]. The Ramachandran plot is presented in picture 7.12. All 123 residues are in the most favoured region. Glu42 was recognized to have a

slightly unusual rotamer (0.7%). The fact that the electron density for the carboxyl group is not extremely clear could have contributed to refine it in a not particular common rotamer. An analysis of the average B-factor of the main chains was also performed with Baverage [143]. The average B-factor is 13 $Å^2$, while the average temperature factor for the mobile areas is higher: 17 $Å^2$ (MH1), 23 $Å^2$ (ML1) 14 $Å^2$ (ML2) and 22 $Å^2$ (ML3). The loop containing the Asp111 has also a higher average B-factor: 18 $Å^2$. The temperature factor analysis and the comparison with the wild type AcpS are shown in Figure 7.13. Again, the B-factor of this mutant is lower than the wild type one, but its pattern is very similar, with the mobile loops characterized by a higher temperature factor.

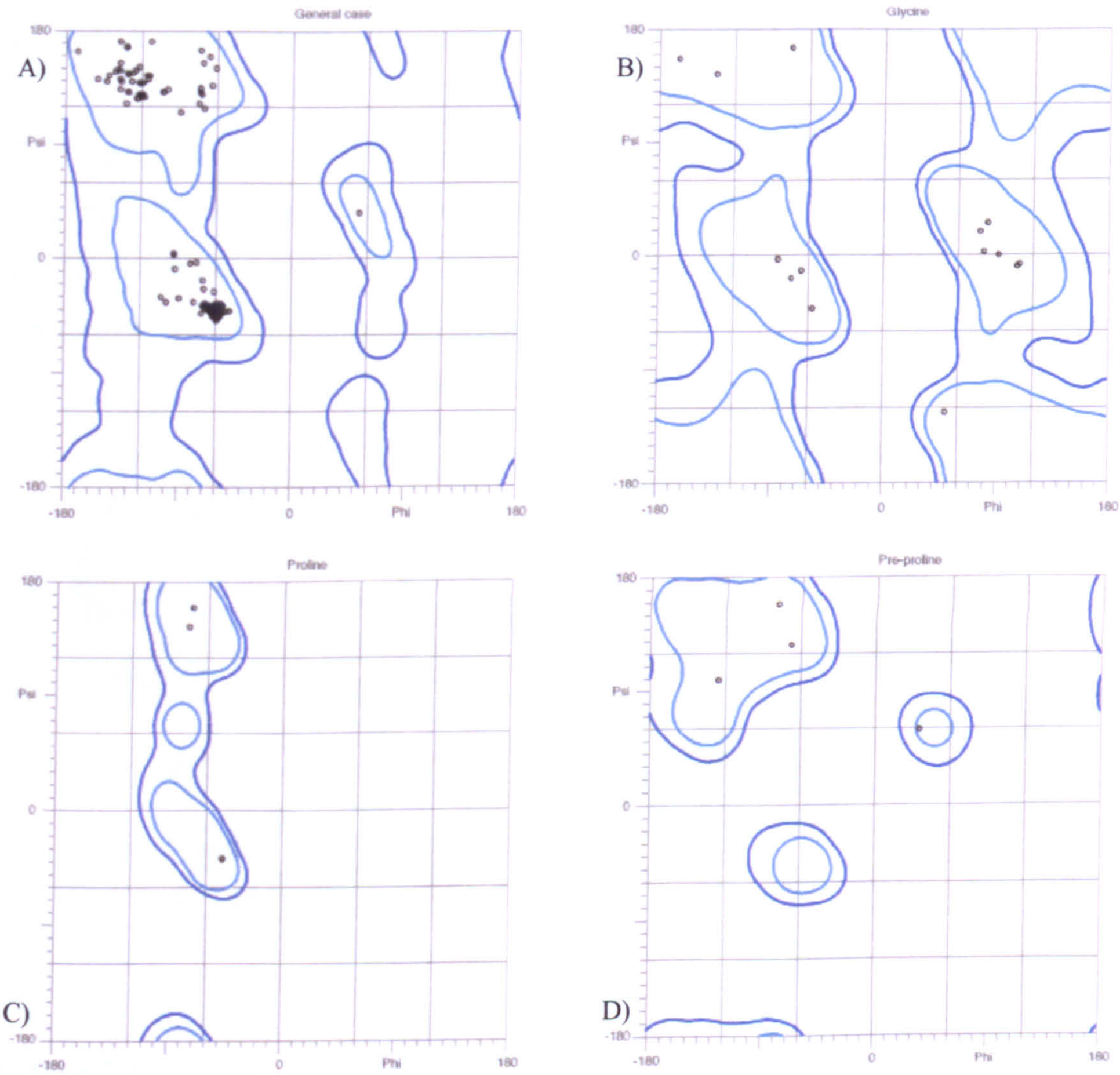


Figure 7.12 Ramachandran plot related to D111A mutant made by MolProbity. A) general case; B) glycine; C) proline; D) pre-proline.

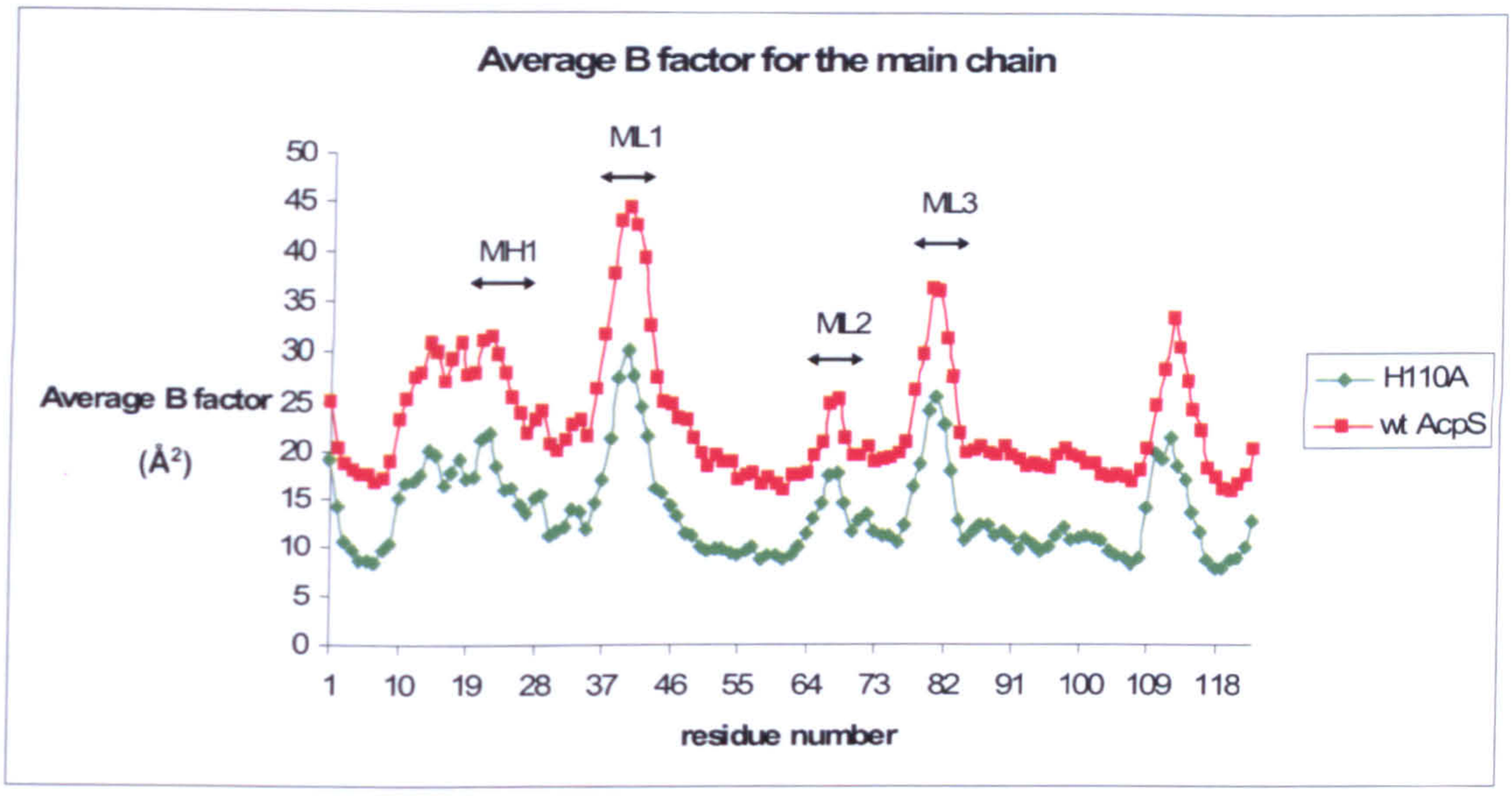


Figure 7.13 H110A B-factor analysis. Average B-factor values for the main chain of H110A mutant and its comparison with the wild type AcpS.

7.7 Analysis of the crystal structure of the D111A AcpS mutant

The mutation of Asp111 to alanine gave a very interesting result. The structure of the D111A mutant contains 2 alternative conformations for the CoA, each of them coordinated to its own magnesium ion (Figure 7.14).

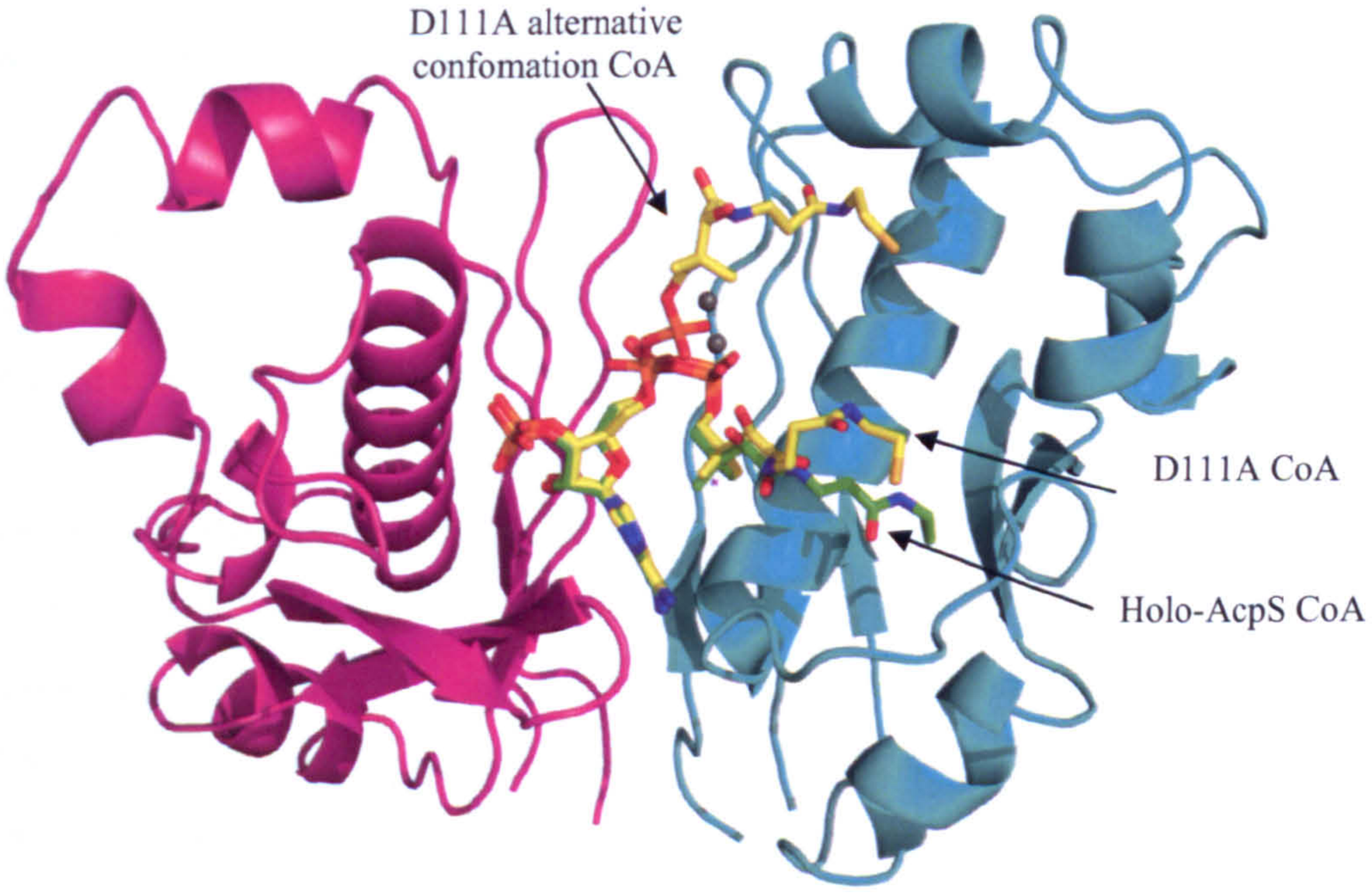


Figure 7.14 D111A AcpS crystal structure. The ribbon representation of D111A AcpS is shown, with monomers A and B coloured in magenta and cyan. CoAs are shown as sticks and coloured according to atom type. CoAs in the D111A AcpS are coloured in yellow, the cofactor present in the holo-AcpS in green. Magnesium ions are shown as grey spheres.

The RMSD after Cα superimposition between the mutant to the wild type is 0.20 Å. The N-terminal region is more disordered and the polyhistidine tag has a different arrangement in comparison with the wild type, explaining the fact that this mutant is eluted at a higher concentration of *imidazole from the nickel* column. Mobile loops are slightly displaced and in the same conformation. Whilst the adenine rings of CoA are placed in the same position, the phosphopantetheine chains are located in a different area. Two possible positions for the phosphopantetheine chain were identified. These orientations were chosen to reflect their best and more likely positions in the model. The first one is similar to the active one in the wild type, with the phosphopantetheine arms that begin to diverge after the carbon C9. After this atom, CoA in the mutant structure seems to assume a more contracted conformation in comparison to a more linear disposition of the chain in the wild type. The alternative CoA lies near sheet β3 until carbon C9, then is surrounded by helices α1 and α2 of another monomer. The alternative phosphopantetheine arm lies on a hydrophobic pocket that is defined by the aliphatic portion of the side chain of Arg15, and by the side chains of Phe16, Leu26, Leu30, Phe53 and Trp71 of a symmetry related subunit. Moreover, additional bits of electron density suggested that the phosphopantetheine tail is mobile, and other locations, especially one that is very similar to the one of CoA in the wild type protein, can also be seen, in particular for the portion after the 2 phosphates. The magnesium coordination sphere seems to be conserved, apart from the fact that some water molecules are missing, probably due to the complexity of this active site and the presence of the 2 alternative conformations of CoA.

7.8 Analysis of the crystal structure of the H110A AcpS mutant

The overall structure of the H110A mutant is very similar to the wild type AcpS, with clear electron density for the alanine that substitutes His110 (Figure 7.15). The RMSD after Cα superimposition between the H110A mutant and the wild type is 0.22 Å. The N-terminal region is more similar to the aspartate mutant than to the wild type AcpS. It is more disordered than the wild type, with a different arrangement. Small displacements can be observed for the mobile loops (approximately between 0.4 Å and 0.7 Å for the Cα portion), which have the same conformation. The cofactor binding site is conserved. The most significant

difference is the orientation of the O6 oxygens of CoA, which are 1.43 Å apart in mutant and wild type structure. The electron density around the β-phosphate suggests that this portion of CoA is a bit more flexible than that one in the wild type structure. The coordination sphere of the magnesium ion is also conserved: the 3 waters, the carboxyl group of Asp9 and the two phosphates of CoA. The side chain of Asp111 is still capable of interaction with a magnesium-coordinating water via hydrogen bonds with another water molecule. Differently to the wild type AcpS, Glu57, residue that is very well conserved in other type I AcpSs, is coordinated to the β-phosphate of CoA, in a conformation that is similar to that one observed in the *B. subtilis* AcpS, where the equivalent residue coordinates the calcium. This residue has been proposed to directly deprotonate the active ACP serine in type II human AcpSs [66].

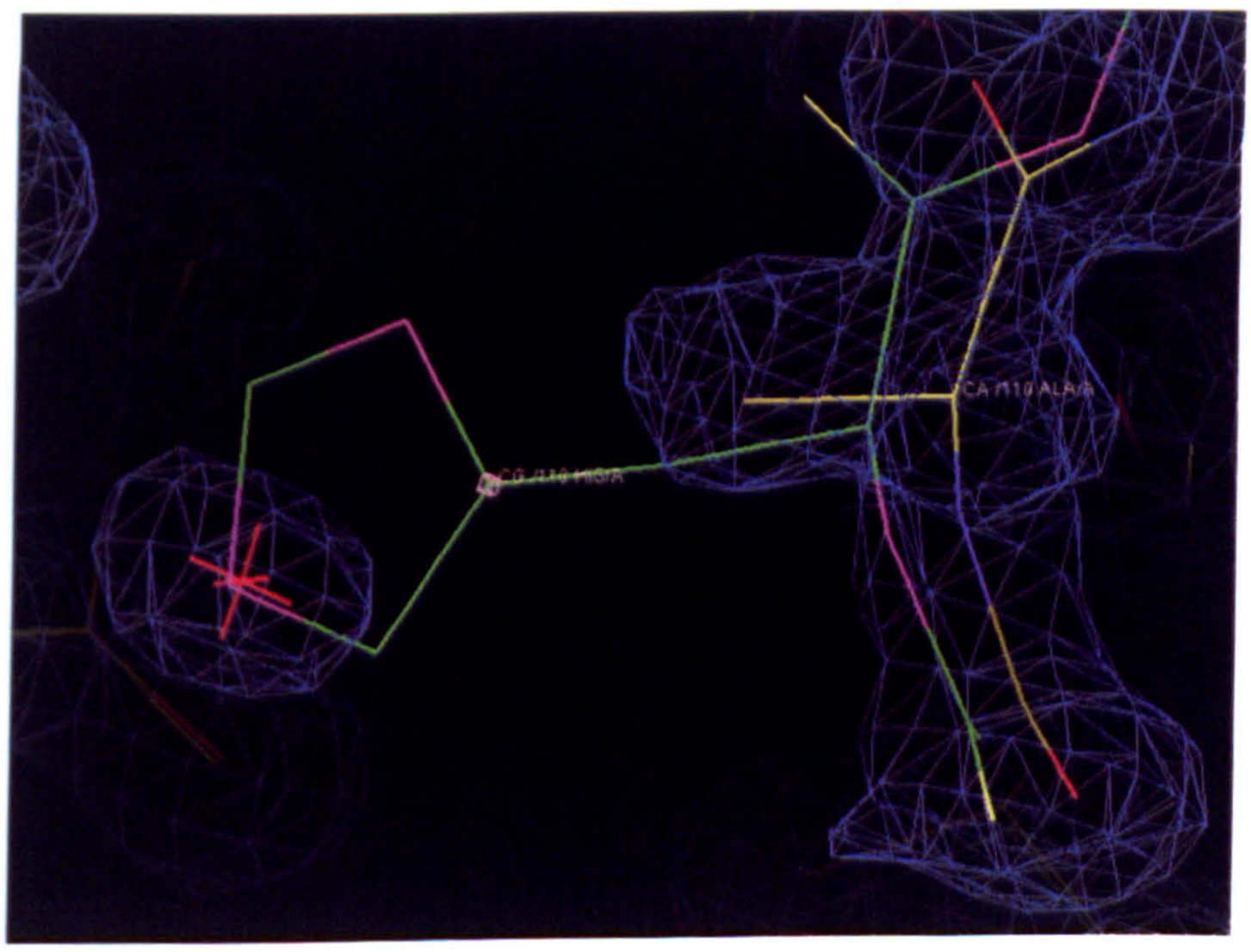


Figure 7.15 The H110A mutation in the corresponding crystal structure. Bonds representation of the alanine in the H110A structure (coloured in yellow) and of the histidine 110 (in green) in the wild type AcpS. The electron density related (FWT, PHWT) to the H110A protein is contoured at 1σ.

7.9 Investigation of CoA binding with ITC

The mutated proteins did not result to be 100% active. In particular, the histidine mutant resulted to be the most affected protein, with a drop of 85% of activity. Whilst an initial low resolution (approximately 4 Å) structure of His110 did not

appear to have CoA present, the highest resolution shows its presence clearly. So, in order for AcpS to phosphopantetheinylate the substrate CoA, the CoA needs to bind, ACP needs to bind in the correct orientation and the active site needs to be organized in such a way that the correct chemistry can be catalyzed. In order to understand the role of substrate binding in the reduced activity of the mutants better, ITC was carried out to investigate the binding of CoA to both the wild type and the mutant enzymes (chapter 3.9). As can be seen from the graphs (Figure 7.16), in the case of the histidine mutant, the affinity for CoA is very similar to the wild type one, and a molar ratio of approximately 1 was calculated (table 7.5). This suggests that the His110 has a more active role in the catalysis or a structural role in the correct formation of the active site rather than in the cofactor binding.

In the D111A crystal structure, 2 alternative conformations of CoAs, and consequently two binding sites, were identified per molecule. This was confirmed by ITC. A strong binding site, with an affinity similar to the wild type protein, was identified after a standard reaction, but the stoichiometry resulted to be either 0.6 or 0.4, suggesting that 1 or 2 monomers per trimer were binding CoA, leaving some subunits available for CoA to occupy an alternative binding site.

	Wild type AcpS	H110A	D111A
K_d (nM)	380 ± 20	450 ± 70	850 ± 40
n	1.2 ± 0.1	0.98 ± 0.05	0.5 ± 0.1
ΔH (cal/mol)	-4000 ± 400	-3500 ± 400	-7000 ± 1800

Table 7.5 Affinity constants and stoichiometry for the H110A and D111A mutants.

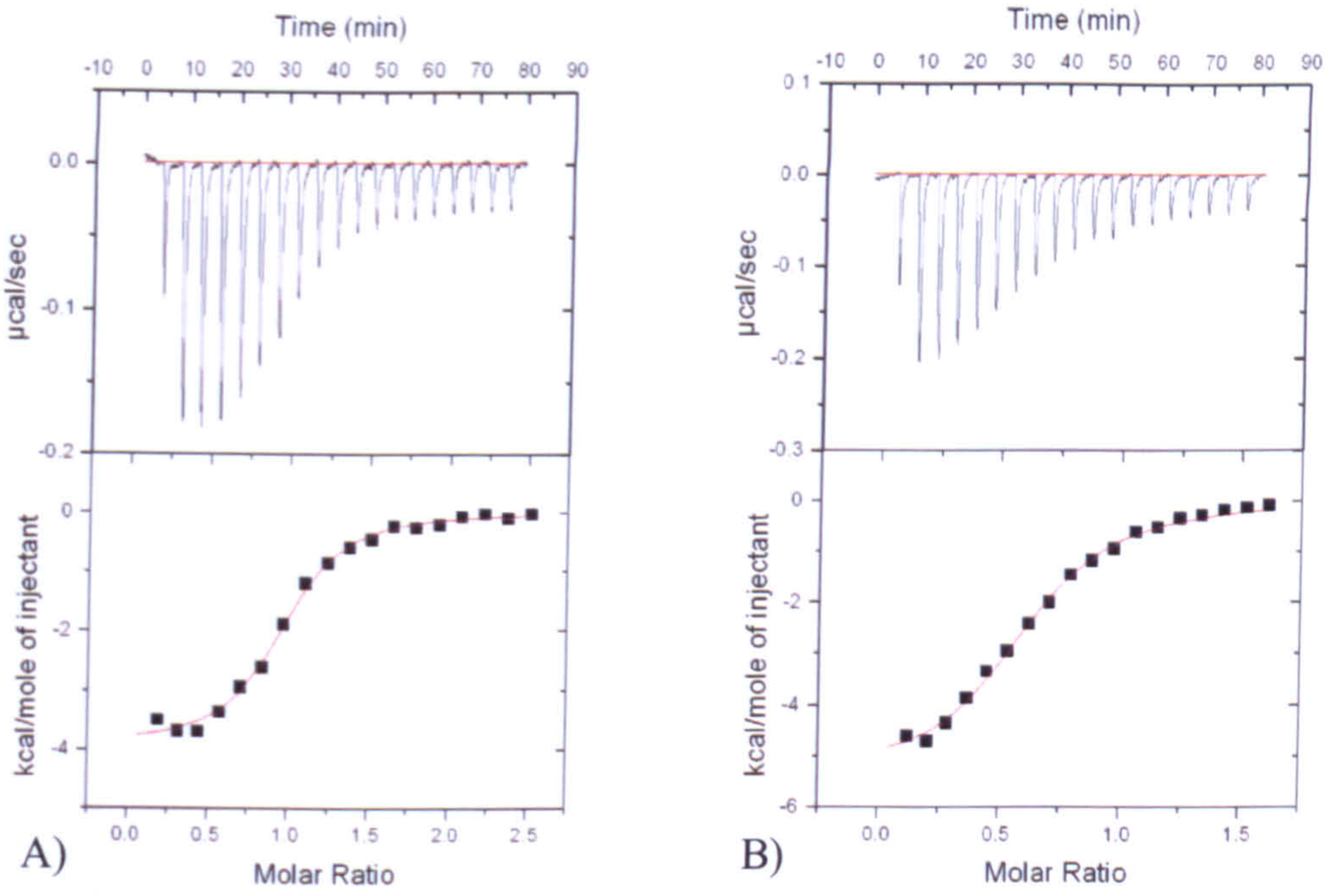


Figure 7.16 ITC graphs showing the presence of a strong binding site for CoA. A) Data related to the H110A mutant; B) data related to the D111A mutant.

So, another experiment with a very high concentration of CoA was carried out, and a very weak binding site could be identified. Firstly, a normal ITC experiment was run (chapter 3.9). This showed an affinity for CoA similar to the wild type, but a stoichiometry of 1:3 or 2:3. So, at the end of this experiment and after complete saturation at that particular CoA concentration, another experiment was carried out with the same protein but with a very high concentration of cofactor (2.5 mM). From this second experiment a weaker binding site was recognized. A blank contained the same concentration of CoA was tested in the absence of the protein, to confirm that the weaker binding site was not a dilution effect. In addition, another experiment was carried out, with a starting CoA concentration of 1.5 mM (Figure 7.17 A). Again, after the first tight binding site and saturation, a weaker site could be identified. Due to the large difference of enthalpy between the two binding sites, it was not possible to determine accurately the K_d for the second binding, but its related graph shows clearly the presence of this weak site. A blank with the wild type protein and with the same volumes of injection was run to check if the peaks related to the weaker binding site were due to the different heat of dilutions (Figure 7.17 B). As expected, the wild type AcpS with a very high concentration of CoA

reaches saturation very quickly, indirectly confirming that the peaks related to the aspartate mutant belong to the weaker site. K_d for this wild type blank was 537 nM, with a stoichiometry of 1.11, values that are very close to those of the wild type AcpS. For the weaker binding site of the aspartate mutant a K_d of 90 μM was calculated, but this value was affected by a big statistical error. The great error that affects the change in enthalpy reflects the fact that 2 different values were calculated: the first for a stoichiometry of 1:3, the other for a 2:3 ratio.

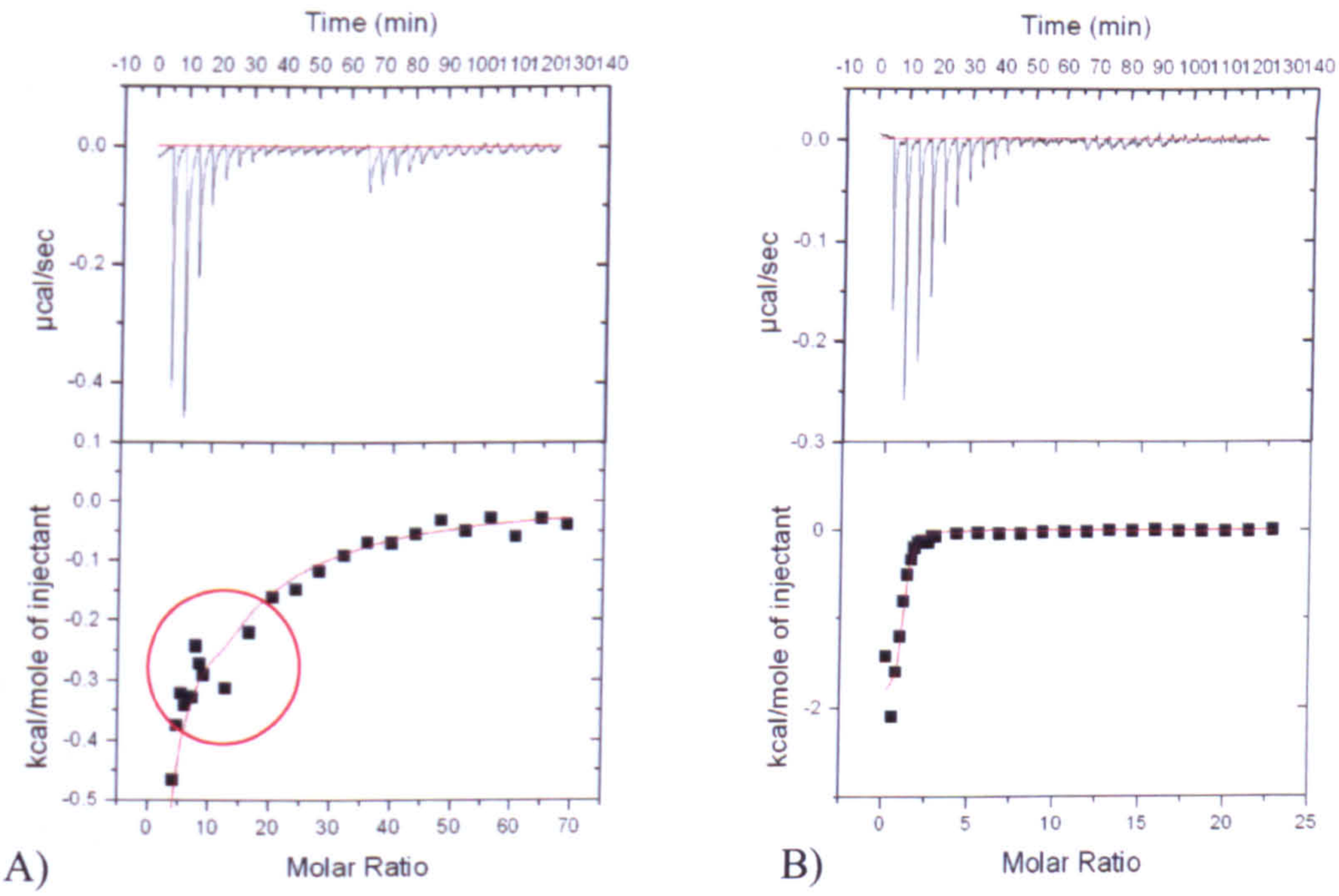


Figure 7.17 Discovering a second binding site for CoA for the aspartate mutant. A) After saturation of the tight binding site, a weaker site is recognized. B) Blank using the wild type AcpS to check if the ΔH characteristic of the second site was due to the heat of dilution.

7.10 Role of the His110 and Asp111 in catalysis

Both His110 and Asp111 are located on either side of CoA, and may be in a position to deprotonate the active serine on ACP. Both mutant enzymes resulted in reduced activity, with the activity of the histidine mutant being most dramatically affected by its mutation to an alanine. ITC confirmed that the CoA binding is preserved. Structurally, for both enzymes only minor perturbations of the protein structure were noticed. The cofactor in the histidine mutant binds essentially in the same way of the cofactor in the wild type AcpS, with the exception that the β-

phosphate of CoA is slightly more variable. In the aspartate mutant, the major difference in comparison with the wild type is in the cofactor binding region, where the electron density shows clearly that the phosphopantetheine chain of CoA can occupy two alternative binding sites. This was confirmed by *ITC experiments* that allowed the identification of a second very weak binding site which may correspond to the second orientation of CoA that was observed in the crystal structure.

7.10.1 Role of His110

From the structural and ITC data it can be proposed that the His110 residue may act as a base in deprotonation of the serine on ACP. The mutation to an alanine almost destroys the activity of the correspondent protein, but the affinity to CoA is not affected. Histidines are commonly present in various active sites and their basic nitrogen of the imidazole ring is used to remove a proton from serine, threonine or cysteine residues to activate them as nucleophiles. The remaining 10% of activity could be explained with the fact that a water molecule is placed at the position of the basic nitrogen of the imidazole ring, and it may act as a weaker base.

Analysis of the structure reveals that a different conformation of the side chain of the histidine can be close either to the serine or to the equivalent of the activating water. If the histidine deprotonates the serine indirectly, an active water will be present in the active site. In order to self hydrolization of CoA to occur, this active water would have to deprotonate another water molecule form the bulk solvent since the distance between the water and the β-phosphate is too great for it to attack it directly.

If His110 is involved as base catalyst, it requires some conformational change to the side chain. This can be promoted by ACP binding, which would explain why the presence of the base in the active site in the absence of ACP does not promote self-hydrolysis.

The observations of Glu57 in an alternative conformation close to that observed for the equivalent residues in *B. subtilis* and human AcpSs rather than coordinating the phosphopantetheine chain, suggests the possibility that this may act as a catalytic base as proposed for the human enzyme.

7.10.2 Role of Asp111

On the contrary, the aspartate 111 seems to have a different structural role during catalysis. D111A AcpS contains an alternative conformation of CoA. The negatively charged oxygens of the aspartate can repulse the negatively charged PO9 oxygen of the tail of the cofactor, and the AO5 and AO6 oxygens of the β-phosphate of CoA, preventing the cofactor from binding in this alternative conformation (Figure 7.18).

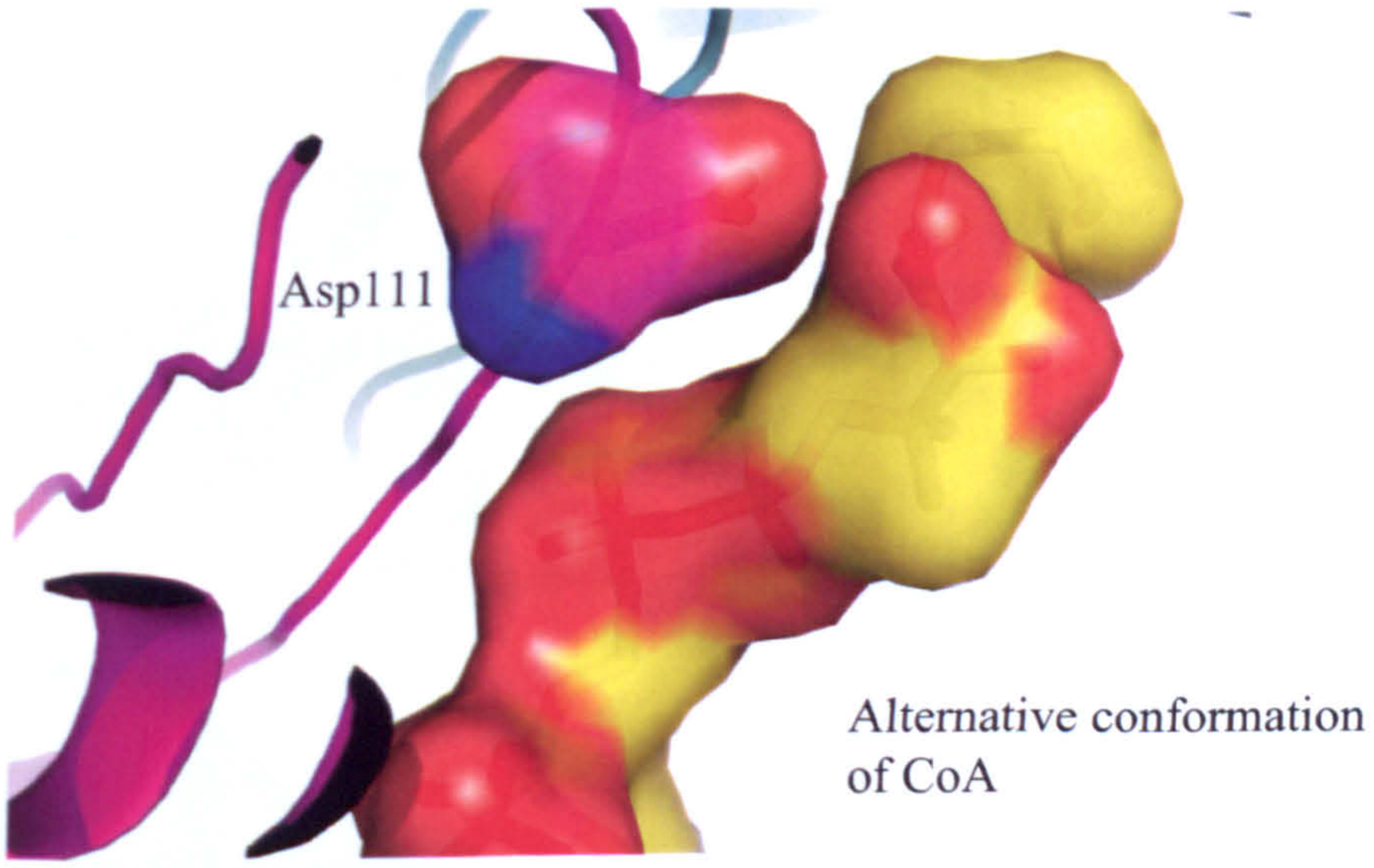


Figure 7.18 The alternative CoA conformation. Charged surface representation of the wild type Asp111 and the alternative CoA in the D111A structure. CoA is shown as sticks and coloured in yellow, according to atom type. AcpS monomer A is coloured in magenta. Positive and negative regions are coloured in blue and red respectively.

The loop containing Asp111 is also very mobile, so it could also be possible that a particular conformation of this region will cause steric clash with the cofactor. Finally, this alternative orientation could also prevent the construction of an active site with the proper organization that allows the correct chemistry to be catalyzed. A model for the binding between the D111A mutant and the ACP was created in the same way as for the wild type (chapter 3.8.7). From Figure 7.19, it is evident that the first carbon atoms after the β-phosphate can clash against the region containing the ACP Asp41. The alternative conformation of the cofactor results to be located between ACP helix α2 and AcpS helix α1, preventing the ACP Asp41

from interacting with AcpS Arg15, and therefore not allowing the interaction that is supposed to be crucial for the beginning of the phosphopantetheinylation reaction. In addition, in this model the active serine clashes against the β-phosphate of CoA. The other conformation of CoA, which is in a similar orientation as the wild type one, seems to permit the phosphopantetheinylation of the ACP, justifying the remaining 55% of activity that characterizes the aspartate mutant.

These results suggest that Asp111 does not have a role as a base in (in)direct deprotonation of Ser42. However, it seems to be necessary in organizing the active site for productive binding of CoA.

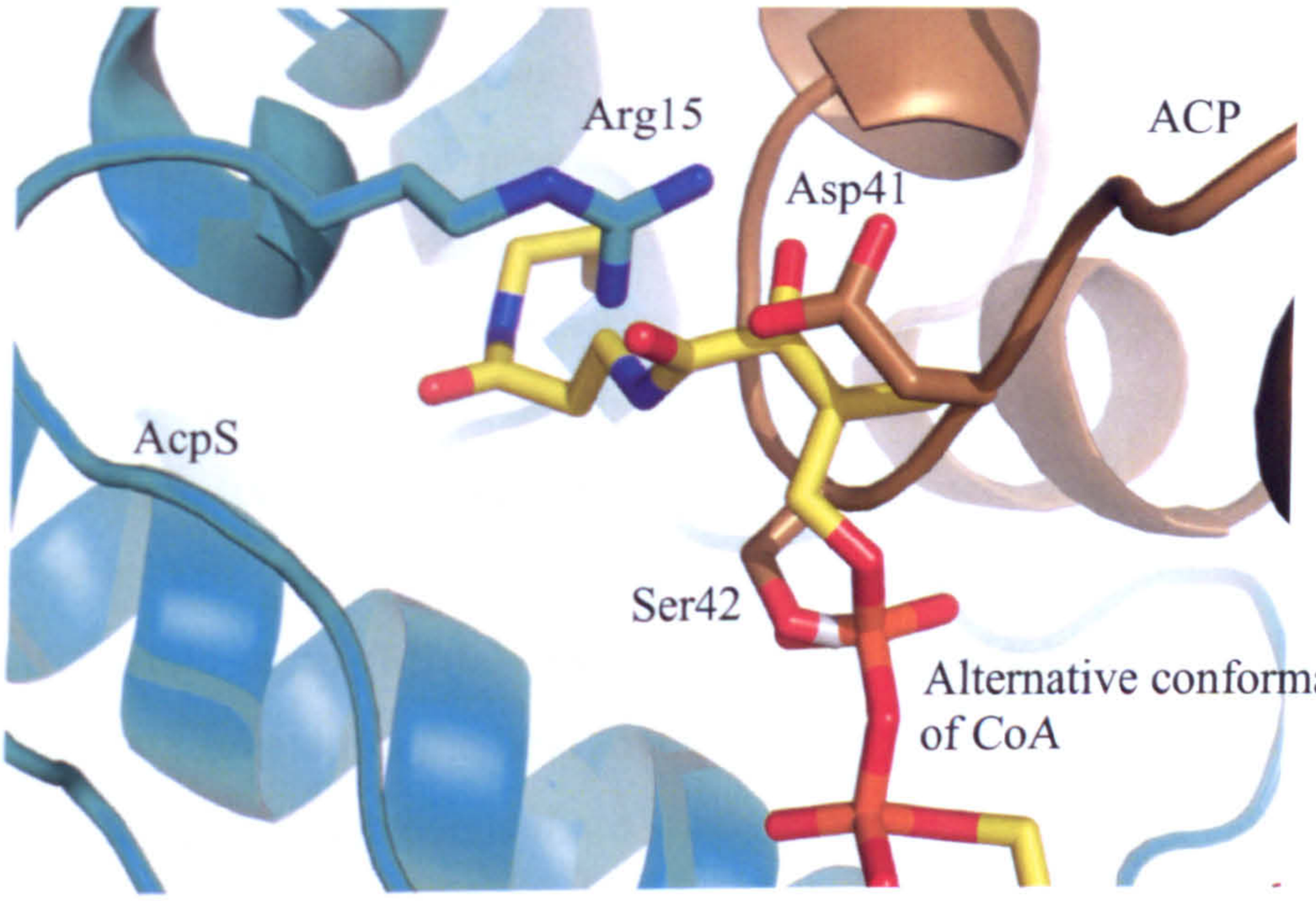


Figure 7.19 Model for ACP binding to D111A mutant. The alternative conformation of CoA could prevent the interaction between ACP Asp41 and AcpS Arg15. The cartoon representation of AcpS is shown in cyan, the modelled ACP in sand. CoA and the other side chains are shown as sticks.

7.10.3 Catalysis revisited

In conclusion, mutagenesis studies of Asp111 suggest a structural role in the binding of CoA and not a catalytic role. The dramatic reduction in activity in the His110 mutant supports a role in catalysis, perhaps as the base in activation of the serine or a water molecule. Analysis of the H110A crystal structure reveals that Glu57, as proposed for the human AcpS, may be directly involved in serine deprotonation. The role of Glu57 was not investigated during the time frame of this thesis.

Conclusions

The work presented in this thesis focuses on the type I actinorhodin acyl carrier protein synthase (AcpS) from *Streptomyces coelicolor*. The cofactor promiscuity and the interaction of AcpS with the acyl carrier protein (ACP), as well as the phosphopantetheinylation mechanism, were investigated.

The crystal structures of the apo-AcpS and of the protein in complex with the cofactor coenzyme A (CoA) were solved at 2 Å and 1.6 Å respectively.

The biological unit is a trimer, with one active site at each interface between 2 subunits. A magnesium ion is present in the electron density in the presence of CoA. This *S. coelicolor* holo-AcpS structure is the first type I AcpS that presents the essential magnesium ion in the active site. The adenine ring of CoA interacts with one monomer, while the phosphopantetheine arm binds to the adjacent monomer. The phosphopantetheine chain lies on a relatively large and well defined hydrophobic pocket.

Four main flexible areas (ML1-3 and MH1) were recognized in the apo-AcpS. These regions are ordered in the holo model but displaced if compared to the apo one. In particular, ML1 surrounds the 3'-phosphate binding pocket, ML2 defines one side of the phosphopantetheine hydrophobic cavity and ML3 stacks against the adenine base of CoA.

By superimposing the apo and the holo protein, it is possible to notice that helix α1 is displaced when CoA is present, indicating a change in the trimer arrangement in the active holo-protein, which results in an "opening up" at the top of the trimer. This displacement could be a signal for the ACP to bind and interact.

Comparing the *S. coelicolor* AcpS to other type I and II AcpSs, it is possible to notice that ML2 is in different conformations. In particular, *S. coelicolor* ML2 seems to be very open, while the equivalent region in the *S. pneumoniae* structure, which is not determined in the presence of 3'-5'-ADP, is very closed, in a conformation similar to the one assumed by the equivalent loop in the *B. subtilis* protein. In the human type II AcpS, the equivalent ML2 is in a intermediate

conformation between the *B. subtilis* and the *S. coelicolor* loops. The high flexibility of ML2 could explain the broader cofactor selection of *S. coelicolor* AcpS. ML2 could enlarge the cavity where the phosphopantetheine chain lies, allowing extended modified CoAs to bind in the active site.

In order to investigate the cofactor promiscuity of AcpS, crystallization trials were performed in the presence of longer acyl-CoAs. The crystal structure of the *S. coelicolor* AcpS in complex with acetyl-CoA was solved at 1.5 Å. This model is very similar to the holo-AcpS. ITC showed that CoA has a higher affinity than the acetyl-CoA for AcpS.

The acetyl-CoA seems to adopt a more contracted conformation rather than causing conformational changes on ML2. All the residues that form the hydrophobic cavity that contains the phosphopantetheine arm are disposed in the same way and with the same conformations. So, the addition of just 3 atoms to the natural cofactor does not seem to cause any dramatic conformational changes.

Several acyl-CoAs were also modelled into the holo-AcpS hydrophobic pocket. The models showed that the pocket was too small to contain these modified CoAs. However, ML2 could assume different conformations in vivo, thus enlarging the phosphopantetheine chain pocket.

To investigate the *S. coelicolor* ACP binding to the AcpS, an AcpS:ACP complex model was made using holo-AcpS and a *S. coelicolor* apo-ACP NMR structure based on a known structure of *B. subtilis* FAS AcpS:ACP complex.

The contacts between the two proteins are predominantly hydrophilic. Interactions between aspartate and glutamate residues on *B. subtilis* ACP helix α3 and arginine residues on *B. subtilis* AcpS helix α1 seem to be conserved in the *S. coelicolor* model. The affinity for apo-AcpS of several ACP mutants in these selected residues is within the same order of magnitude, suggesting that, even though the hydrophilic interactions are important for the interaction with the AcpS and are necessary to tune the precise location of ACP in the AcpS active site to allow catalysis, other hydrophobic interactions are also involved and participate to lock the ACP in the right position. The displacement of AcpS helix α1 and MH1 after CoA binding could also be a signal for ACP that AcpS is ready for catalysis.

The mutagenesis of the conserved ACP Asp41 into an alanine results in a significant change to binding affinity, and in complete inactivation of the phosphopantetheine transfer. The mutations of Asp41 to a glutamate affects the affinity only partially, and results in a 85% loss of activity. So, even if Asp41 is not involved in any chemical reaction, it could be important for locating the loop containing the activating ACP serine in the right position.

After structural analysis of the AcpS:ACP model, it was proposed that the interaction between Asp41 in ACP and Arg15 in AcpS could lock ACP and more importantly the Ser42 into the right position. *Asp41 is located in a region* on ACP helix α2 that can undergo a conformational change, permitting the formation of the salt bond between Arg15 and Asp41, thus bringing the Ser42 near the active site in a position where it could attack the β-phosphate from either side of CoA. Ser42 can replace water Z183, which coordinates with Mg^{2+} and is also within hydrogen bonding distance from AcpS Asp111, or water Z172 which is potentially within hydrogen bonding distance from AcpS His110. After a small conformational change, the aspartate or the histidine residues may deprotonate the hydroxyl group of Ser42 directly or indirectly via an activating water. If an activating water is involved, the coordinated magnesium will have a catalytic role, in addition to its role in binding CoA.

For the *S. coelicolor* AcpS, a preformed active site might be partially responsible for the promiscuity. These residues could be *in situ* as base to assist the reaction, without the need of an additional base contributed by ACP which would require precise positioning.

AcpS mutants H110A and D111A were characterized biochemically and their crystal structures in complex with CoA were solved at 1.3 Å and 1.4 Å respectively.

D111A AcpS resulted with a 45% reduction in activity, while in the H110A AcpS the activity was reduced by approximately 85%. ITC was carried out to determine the cofactor binding constant. The affinity for CoA of the histidine mutant is very similar to the wild type one, suggesting that His110 has a more active role in the catalysis or a structural role in the correct formation of the active site rather than in the cofactor binding.

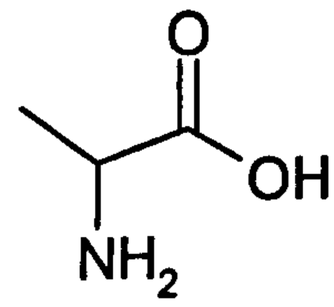
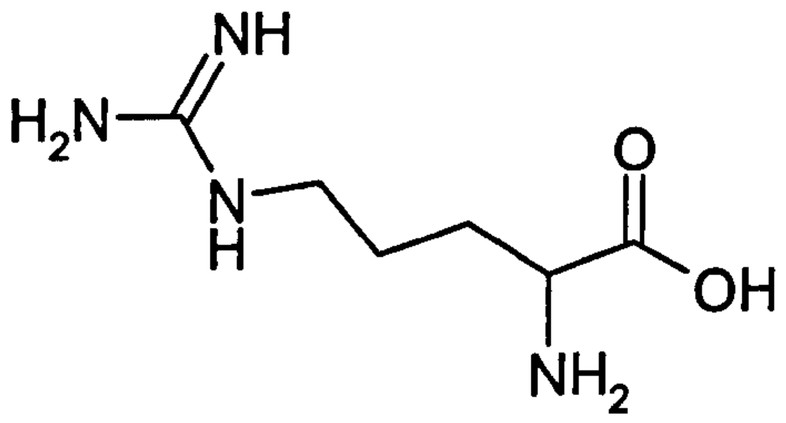
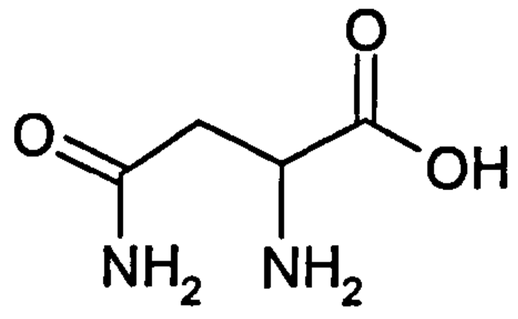
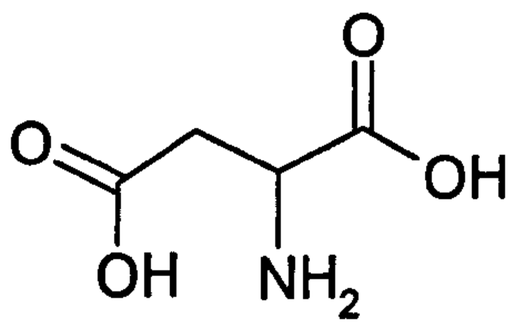
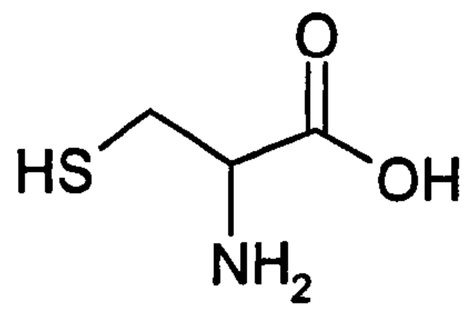
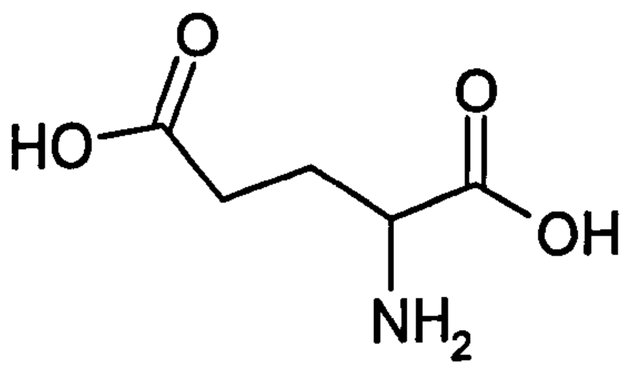
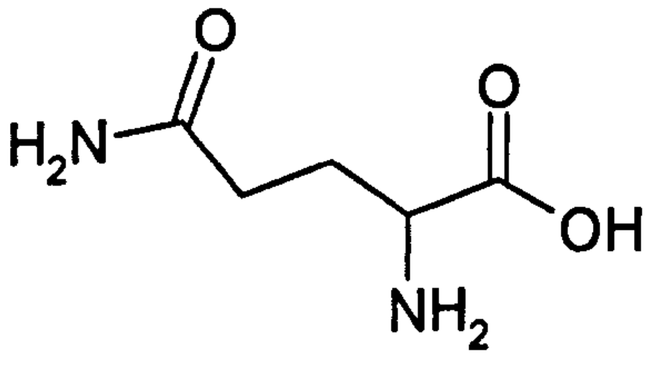
Conclusions

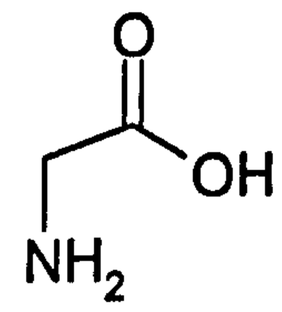
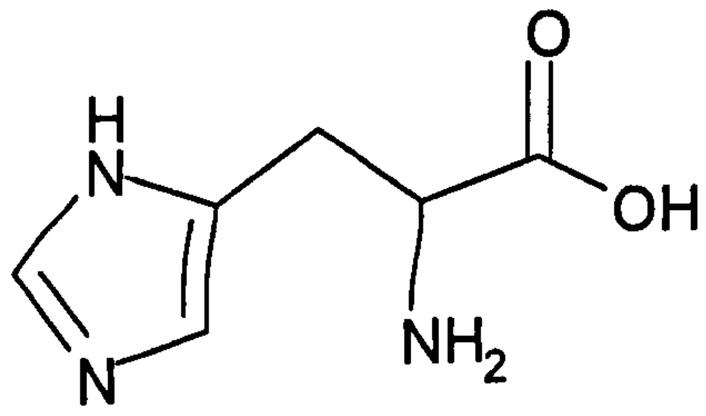
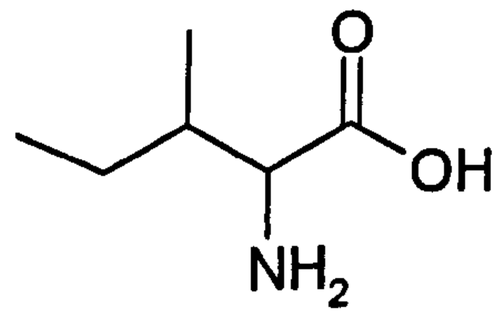
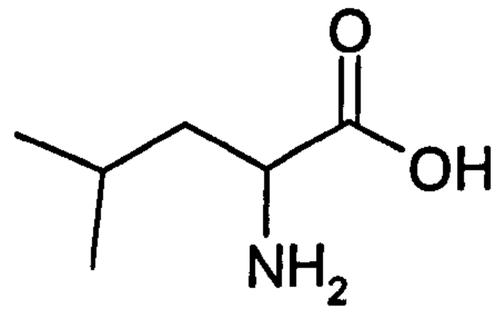
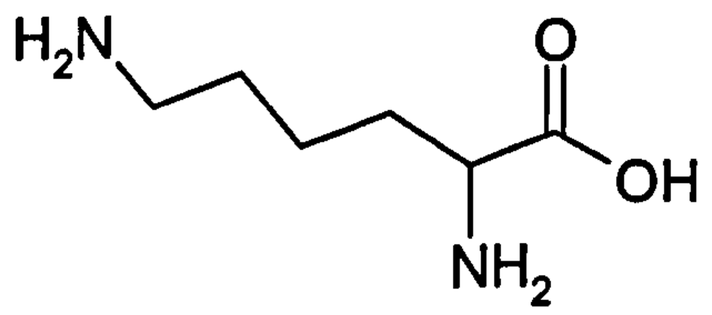
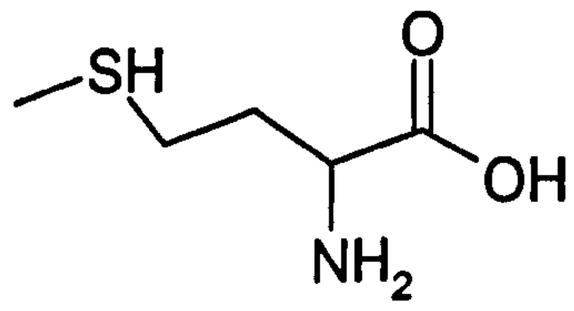
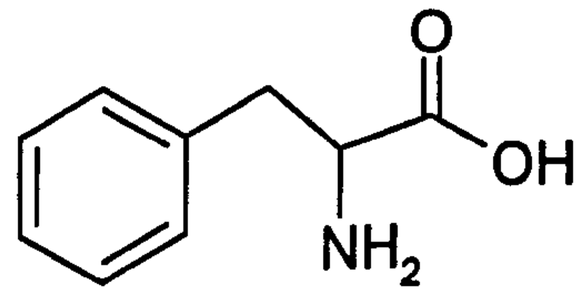
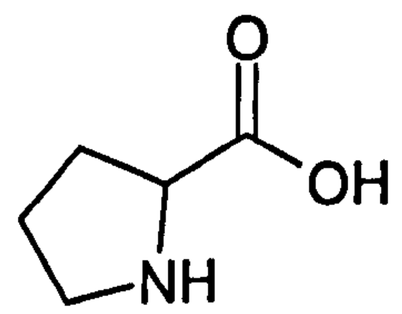
These experiments and analysis of the crystal structure H110A AcpS suggest that His110 could deprotonate the ACP serine directly or indirectly via an activating water.

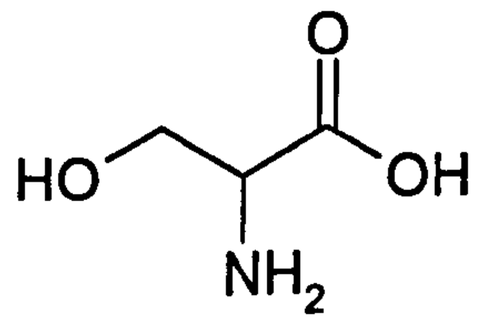
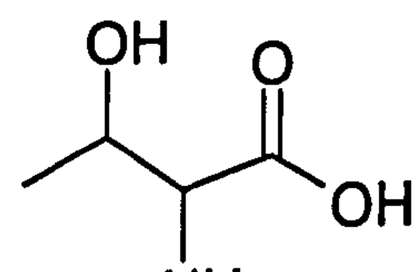
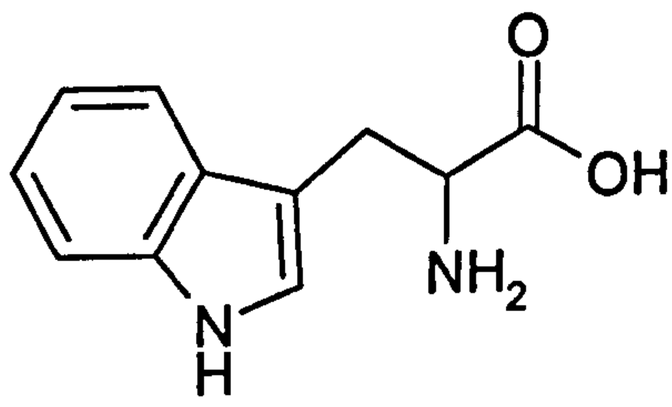
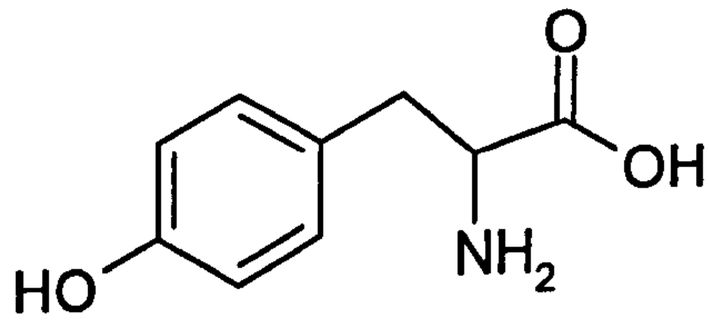
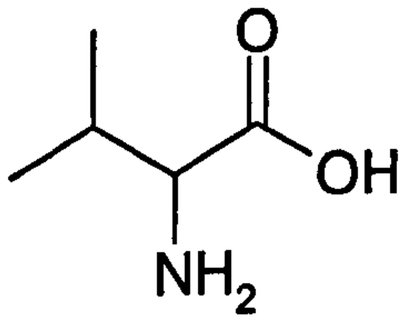
On the contrary, Asp111 seems to have a different structural role during catalysis. Inspection of the electron density at the CoA binding site and ITC experiments suggested that CoA was bound in two alternative conformations in the D111A AcpS mutant, one in a similar way as in the holo-AcpS, and a second one that adopts a completely new conformation and lies on a different hydrophobic pocket. The negatively charged oxygens of the aspartate could repulse the negatively charged oxygens of CoA, preventing the cofactor from binding in this alternative conformation. In addition, the loop containing Asp111 is also very mobile, so it could also be possible that a particular conformation of this region causes steric clash with the cofactor. So, these experiments suggest a structural role in the binding of CoA and not a catalytic role.

In conclusion, the structure of AcpS from *S. coelicolor* has been determined in a catalytically competent complex which reveals a different location for the β-phosphate in the active site compared to that observed in *B. subtilis* type I AcpS. In contrast to the model proposed for the *B. subtilis* type I AcpS, where a water molecule activated by another water or by a base on ACP is thought to deprotonate the ACP serine, in the promiscuous *S. coelicolor* AcpS a residue is proposed to deprotonate (in)directly the active serine on ACP, avoiding the precise positioning of a deprotonating residue located on ACP. This is similar to what has been proposed for the human type II AcpS.

Appendix 1 Amino Acids

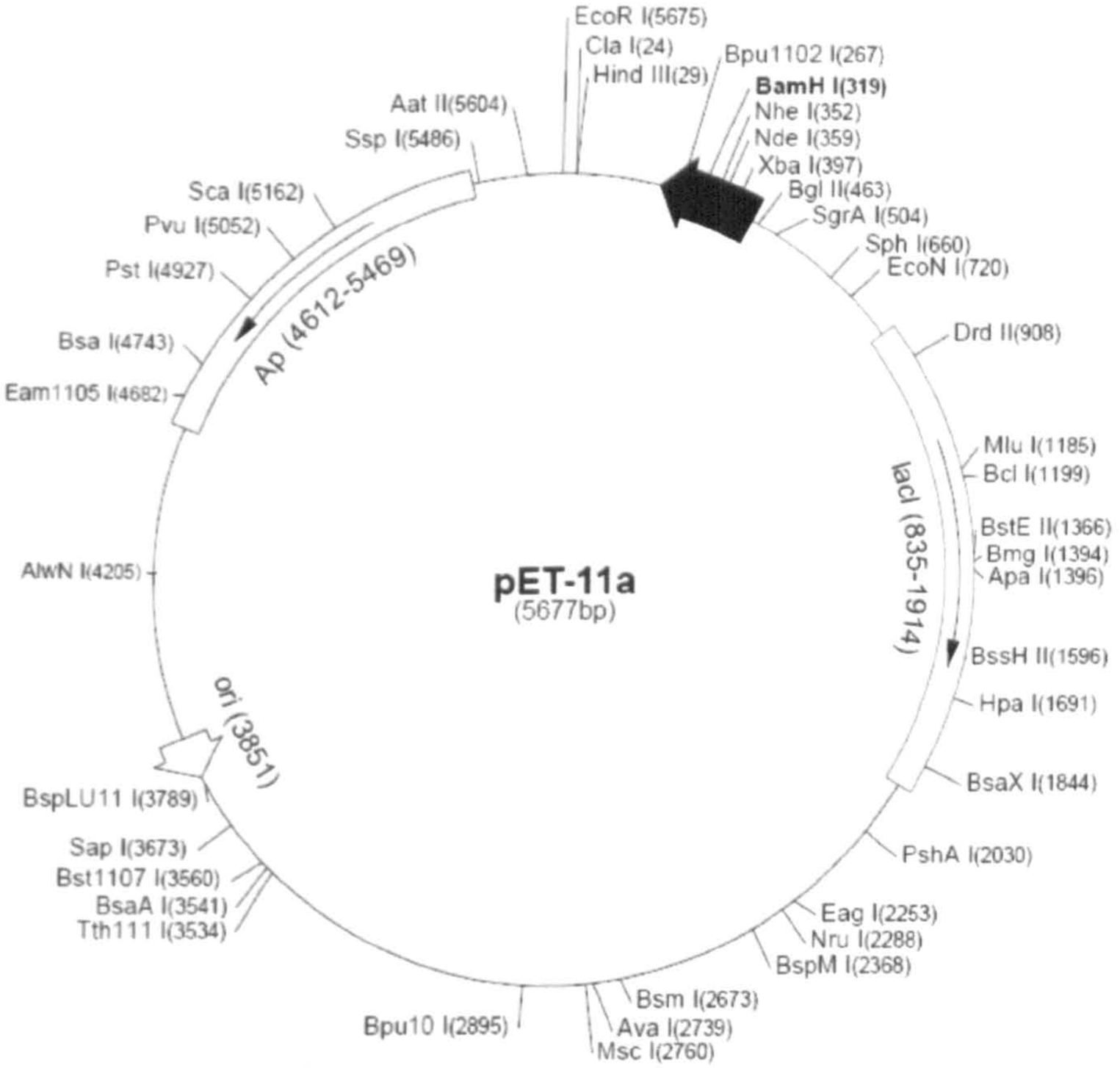
Amino acid	3 letter code	1 letter code	Structure
Alanine	Ala	A	
Arginine	Arg	R	
Asparagine	Asn	N	
Aspartate	Asp	D	
Cysteine	Cys	C	
Glutamate	Glu	E	
Glutamine	Gln	Q	

Glycine	Gly	G	
Histidine	His	H	
Isoleucine	Ile	I	
Leucine	Leu	L	
Lysine	Lys	K	
Methionine	Met	M	
Phenylalanine	Phe	F	
Proline	Pro	P	

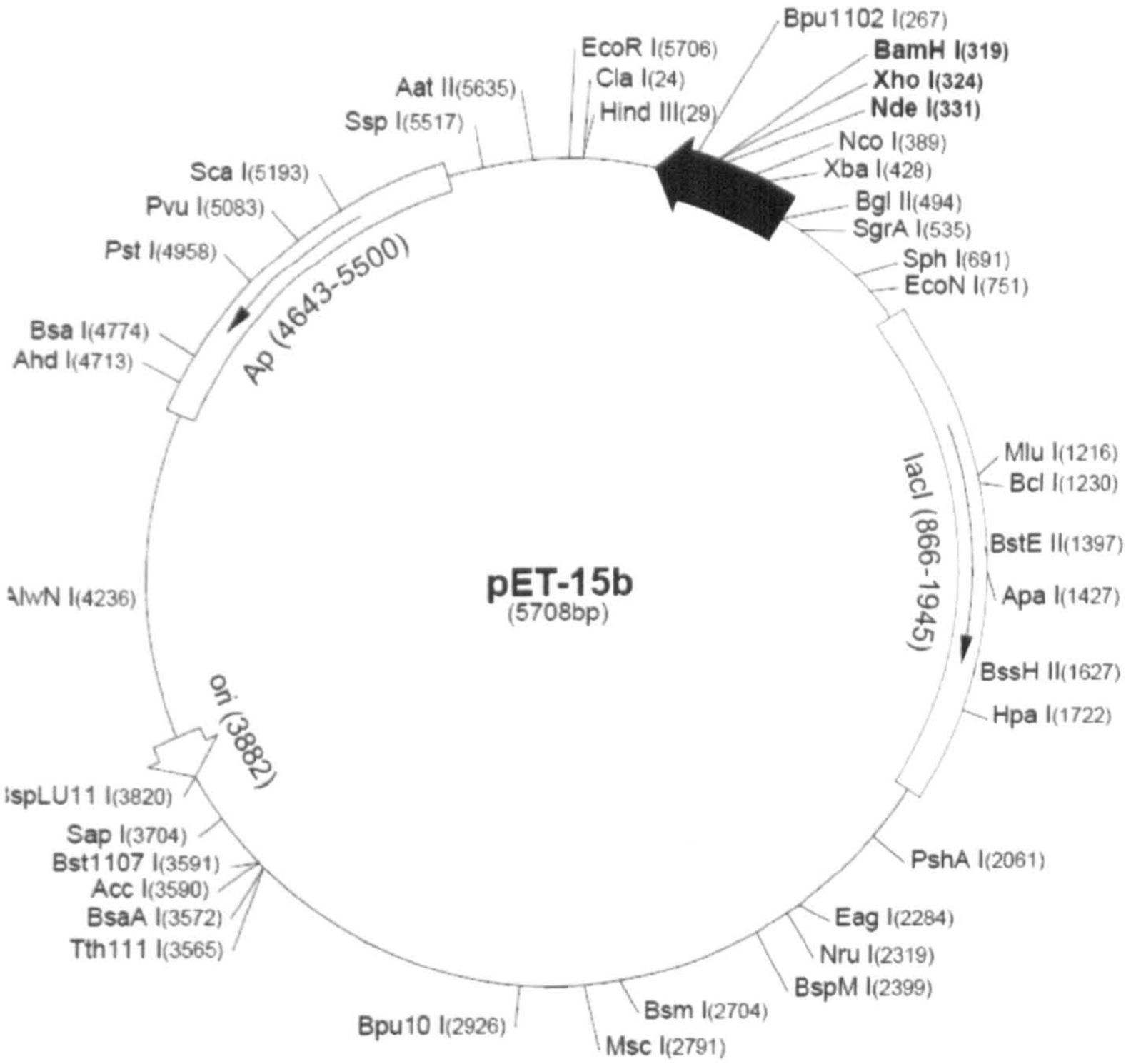
Serine	Ser	S	
Threonine	Thr	T	
Tryptophan	Trp	W	
Tyrosin	Tyr	T	
Valine	Val	V	

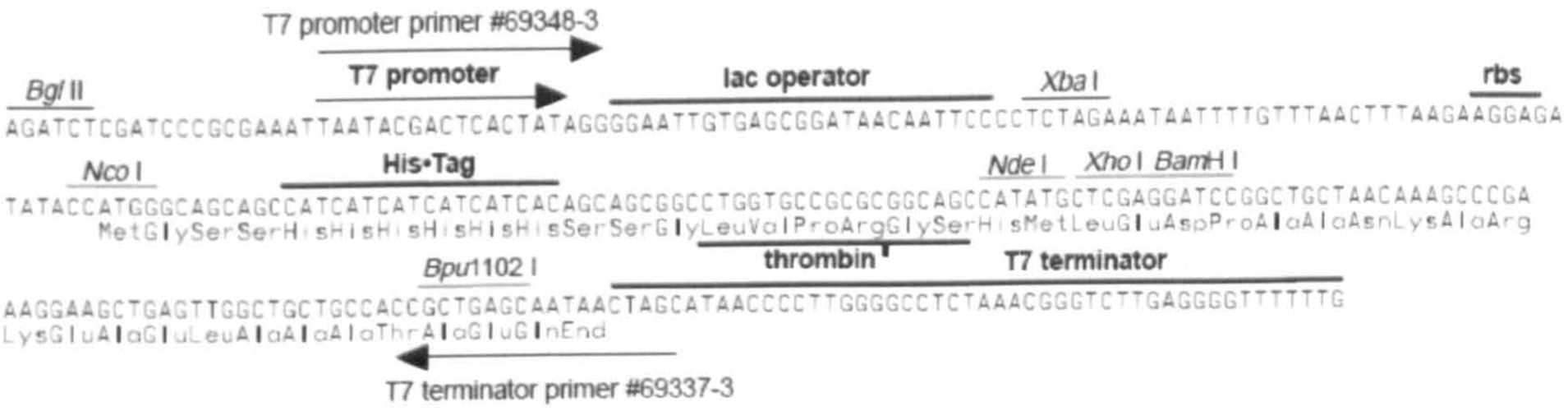
Appendix 2 Vector Maps

pET-11a-d



pET-15b


pET-15b
(5708bp)
Bpu1102 I(267)
BamH I(319)
Xho I(324)
Nde I(331)
Nco I(389)
Xba I(428)
Bgl II(494)
SgrA I(535)
Sph I(691)
EcoN I(751)
EcoR I(5706)
Cla I(24)
Hind III(29)
Aat II(5635)
Ssp I(5517)
Sca I(5193)
Pvu I(5083)
Pst I(4958)
Bsa I(4774)
Ahd I(4713)
Ap (4643-5500)
lacI (866-1945)
Mlu I(1216)
Bcl I(1230)
BstE II(1397)
Apa I(1427)
BssH II(1627)
Hpa I(1722)
PshA I(2061)
Eag I(2284)
Nru I(2319)
BspM I(2399)
Bsm I(2704)
Msc I(2791)
Bpu10 I(2926)
AlwN I(4236)
ori (3882)
BspLU11 I(3820)
Sap I(3704)
Bst1107 I(3591)
Acc I(3590)
BsaA I(3572)
Tth111 I(3565)


T7 promoter primer #69348-3
T7 promoter
lac operator
Bgl II
Xba I
rbs
AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGA
Nco I
His•Tag
Nde I
Xho I
BamH I
TATACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTCGAGGATCCGGCTGCTAACAAAGCCCGA
MetGlySerSerHisHisHisHisHisHisSerSerGlyLeuValProArgGlySerHisMetLeuGluAspProAlaAlaAsnLysAlaArg
thrombin
T7 terminator
Bpu1102 I
AAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG
LysGluAlaGluLeuAlaAlaAlaThrAlaGluGlnEnd
T7 terminator primer #69337-3

Appendix 3 Crystallization kits

Crystal Screens 1 and 2 (Hampton Research)

Tube	Salt	Buffer	Precipitant
1	0.02 M Calcium chloride dihydrate	0.1 M Sodium acetate trihydrate pH 4.6	30% v/v (+/-)-2-Methyl-2,4-pentanediol
2	None	None	0.4 M Potassium sodium tartrate tetrahydrate
3	None	None	0.4 M Ammonium phosphate monobasic
4	None	0.1 M Tris hydrochloride pH 8 5	2.0 M Ammonium sulfate
5	0.2 M Sodium citrate tribasic dihydrate	0.1 M HEPES sodium pH 7.5	30% v/v (+/-)-2-Methyl-2,4-pentanediol
6	0.2 M Magnesium chloride hexahydrate	0.1 M Tris hydrochloride pH 8.5	30% w/v Polyethylene glycol 4,000
7	None	0.1 M Sodium cacodylate trihydrate pH 6.5	1.4 M Sodium acetate trihydrate
8	0.2 M Sodium citrate tribasic dihydrate	0.1 M Sodium cacodylate trihydrate pH 6.5	30% v/v 2-Propanol
9	0.2 M Ammonium acetate	0.1 M Sodium citrate tribasic dihydrate pH 5.6	30% w/v Polyethylene glycol 4,000
10	0.2 M Ammonium acetate	0.1 M Sodium acetate trihydrate pH 4.6	30% w/v Polyethylene glycol 4,000
11	None	0.1 M Sodium citrate tribasic dihydrate pH 5.6	1.0 M Ammonium phosphate monobasic
12	0.2 M Magnesium chloride hexahydrate	0.1 M HEPES sodium pH 7.5	30% v/v 2-Propanol
13	0.2 M Sodium citrate tribasic dihydrate	0.1 M Tris hydrochloride pH 8.5	30% v/v Polyethylene glycol 400
14	0.2 M Calcium chloride dihydrate	0.1 M HEPES sodium pH 7.5	28% v/v Polyethylene glycol 400
15	0.2 M Ammonium sulfate	0.1 M Sodium cacodylate trihydrate pH 6.5	30% w/v Polyethylene glycol 8,000
16	None	0.1 M HEPES sodium pH 7.5	1.5 M Lithium sulfate monohydrate
17	0.2 M Lithium sulfate monohydrate	0.1 M Tris hydrochloride pH 8.5	30% w/v Polyethylene glycol 4,000
18	0.2 M Magnesium acetate tetrahydrate	0.1 M Sodium cacodylate trihydrate pH 6.5	20% w/v Polyethylene glycol 8,000
19	0.2 M Ammonium acetate	0.1 M Tris hydrochloride pH 8.5	30% v/v 2-Propanol
20	0 2 M Ammonium sulfate	0.1 M Sodium acetate trihydrate pH 4 6	25% w/v Polyethylene glycol 4,000
21	0.2 M Magnesium acetate tetrahydrate	0.1 M Sodium cacodylate trihydrate pH 6.5	30% v/v (+/-)-2-Methyl-2,4-pentanediol
22	0.2 M Sodium acetate trihydrate	0.1 M Tris hydrochloride pH 8.5	30% w/v Polyethylene glycol 4,000
23	0.2 M Magnesium chloride hexahydrate	0.1 M HEPES sodium pH 7.5	30% v/v Polyethylene glycol 400
24	0.2 M Calcium chloride dihydrate	0.1 M Sodium acetate trihydrate pH 4.6	20% v/v 2-Propanol
25	None	0.1 M Imidazole pH 6.5	1.0 M Sodium acetate trihydrate
26	0.2 M Ammonium acetate	0.1 M Sodium citrate tribasic dihydrate pH 5.6	30% v/v (+/-)-2-Methyl-2,4-pentanediol
27	0.2 M Sodium citrate tribasic dihydrate	0.1 M HEPES sodium pH 7.5	20% v/v 2-Propanol
28	0.2 M Sodium acetate trihydrate	0.1 M Sodium cacodylate trihydrate pH 6 5	30% w/v Polyethylene glycol 8,000
29	None	0.1 M HEPES sodium pH 7.5	0.8 M Potassium sodium tartrate tetrahydrate
30	0.2 M Ammonium sulfate	None	30% w/v Polyethylene glycol 8,000
31	0.2 M Ammonium sulfate	None	30% w/v Polyethylene glycol 4,000
32	None	None	2.0 M Ammonium sulfate
33	None	None	4.0 M Sodium formate
34	None	0.1 M Sodium acetate trihydrate pH 4 6	2.0 M Sodium formate
35	None	0.1 M HEPES sodium pH 7.5	0.8 M Sodium phosphate monobasic monohydrate 0.8 M Potassium phosphate monobasic
36	None	0.1 M Tris hydrochloride pH 8.5	8% w/v Polyethylene glycol 8,000
37 38	None None	0.1 M Sodium acetate trihydrate pH 4.6 0.1 M HEPES sodium pH 7.5	8% w/v Polyethylene glycol 4,000 1.4 M Sodium citrate tribasic dihydrate
39	None	0.1 M HEPES sodium pH 7.5	2% v/v Polyethylene glycol 400 2.0 M Ammonium sulfate
40	None	0.1 M Sodium citrate tribasic dihydrate pH 5.6	20% v/v 2-Propanol 20% w/v Polyethylene glycol 4,000
41	None	0.1 M HEPES sodium pH 7.5	10% v/v 2-Propanol 20% w/v Polyethylene glycol 4,000
42	0.05 M Potassium phosphate monobasic	None	20% w/v Polyethylene glycol 8,000
43	None	None	30% w/v Polyethylene glycol 1,500
44	None	None	0.2 M Magnesium formate dihydrate
45	0.2 M Zinc acetate dihydrate	0.1 M Sodium cacodylate trihydrate pH 6.5	18% w/v Polyethylene glycol 8,000
46	0.2 M Calcium acetate hydrate	0.1 M Sodium cacodylate trihydrate pH 6.5	18% w/v Polyethylene glycol 8,000
47	None	0.1 M Sodium acetate trihydrate pH 4 6	2.0 M Ammonium sulfate
48	None	0.1 M Tris hydrochloride pH 8.5	2.0 M Ammonium phosphate monobasic
49	1.0 M Lithium sulfate monohydrate	None	2% w/v Polyethylene glycol 8,000
50	0.5 M Lithium sulfate monohydrate	None	15% w/v Polyethylene glycol 8,000

PEG/Ion Screen (Hampton Research)

Tube	Salt	Polymer	pH
1	0.2 M Sodium fl uoride	20% w/v Polyethylene glycol 3,350	7.3
2	0.2 M Potassium fl uoride	20% w/v Polyethylene glycol 3,350	7.3
3	0.2 M Ammonium fl uoride	20% w/v Polyethylene glycol 3,350	6.2
4	0.2 M Lithium chloride	20% w/v Polyethylene glycol 3,350	6.8
5	0.2 M Magnesium chloride hexahydrate	20% w/v Polyethylene glycol 3,350	5.9
6	0.2 M Sodium chloride	20% w/v Polyethylene glycol 3,350	6.9
7	0.2 M Calcium chloride dihydrate	20% w/v Polyethylene glycol 3,350	5.1
8	0.2 M Potassium chloride	20% w/v Polyethylene glycol 3,350	7.0
9	0.2 M Ammonium chloride	20% w/v Polyethylene glycol 3,350	6.3
10	0.2 M Sodium iodide	20% w/v Polyethylene glycol 3,350	7.0
11	0.2 M Potassium iodide	20% w/v Polyethylene glycol 3,350	7.0
12	0.2 M Ammonium iodide	20% w/v Polyethylene glycol 3,350	6.2
13	0.2 M Sodium thiocyanate	20% w/v Polyethylene glycol 3,350	6.9
14	0.2 M Potassium thiocyanate	20% w/v Polyethylene glycol 3,350	7.0
15	0.2 M Lithium nitrate	20% w/v Polyethylene glycol 3,350	7.1
16	0.2 M Magnesium nitrate hexahydrate	20% w/v Polyethylene glycol 3,350	5.9
17	0.2 M Sodium nitrate	20% w/v Polyethylene glycol 3,350	6.8
18	0.2 M Potassium nitrate	20% w/v Polyethylene glycol 3,350	6.8
19	0.2 M Ammonium nitrate	20% w/v Polyethylene glycol 3,350	6.2
20	0.2 M Magnesium formate dihydrate	20% w/v Polyethylene glycol 3,350	7.0
21	0.2 M Sodium formate	20% w/v Polyethylene glycol 3,350	7.2
22	0.2 M Potassium formate	20% w/v Polyethylene glycol 3,350	7.3
23	0.2 M Ammonium formate	20% w/v Polyethylene glycol 3,350	6.6
24	0.2 M Lithium acetate dihydrate	20% w/v Polyethylene glycol 3,350	7.9
25	0.2 M Magnesium acetate tetrahydrate	20% w/v Polyethylene glycol 3,350	7.9
26	0.2 M Zinc acetate dihydrate	20% w/v Polyethylene glycol 3,350	6.4
27	0.2 M Sodium acetate trihydrate	20% w/v Polyethylene glycol 3,350	8.0
28	0.2 M Calcium acetate hydrate	20% w/v Polyethylene glycol 3,350	7.5
29	0.2 M Potassium acetate	20% w/v Polyethylene glycol 3,350	8.1
30	0.2 M Ammonium acetate	20% w/v Polyethylene glycol 3,350	7.1
31	0.2 M Lithium sulfate monohydrate	20% w/v Polyethylene glycol 3,350	6.0
32	0.2 M Magnesium sulfate heptahydrate	20% w/v Polyethylene glycol 3,350	6.0
33	0.2 M Sodium sulfate decahydrate	20% w/v Polyethylene glycol 3,350	6.7
34	0.2 M Potassium sulfate	20% w/v Polyethylene glycol 3,350	6.8
35	0.2 M Ammonium sulfate	20% w/v Polyethylene glycol 3,350	6.0
36	0.2 M Sodium tartrate dibasic dihydrate	20% w/v Polyethylene glycol 3,350	7.3
37	0.2 M Potassium sodium tartrate tetrahydrate	20% w/v Polyethylene glycol 3,350	7.4
38	0.2 M Ammonium tartrate dibasic	20% w/v Polyethylene glycol 3,350	6.6
39	0.2 M Sodium phosphate monobasic monohydrate	20% w/v Polyethylene glycol 3,350	4.7
40	0.2 M Sodium phosphate dibasic dihydrate	20% w/v Polyethylene glycol 3,350	9.1
41	0.2 M Potassium phosphate monobasic	20% w/v Polyethylene glycol 3,350	4.8
42	0.2 M Potassium phosphate dibasic	20% w/v Polyethylene glycol 3,350	9.2
43	0.2 M Ammonium phosphate monobasic	20% w/v Polyethylene glycol 3,350	4.6
44	0.2 M Ammonium phosphate dibasic	20% w/v Polyethylene glycol 3,350	8.0
45	0.2 M Lithium citrate tribasic tetrahydrate	20% w/v Polyethylene glycol 3,350	8.4
46	0.2 M Sodium citrate tribasic dihydrate	20% w/v Polyethylene glycol 3,350	8.3
47	0.2 M Potassium citrate tribasic monohydrate	20% w/v Polyethylene glycol 3,350	8.3
48	0.2 M Ammonium citrate dibasic	20% w/v Polyethylene glycol 3,350	5.1

Clear Strategy Screen I (MDL)

1 0.3 M NaAcetate 25% PEG 2K MME	2 0.2 M Li_2SO_4 25% PEG 2K MME	3 0.2 M $MgCl_2$ 25% PEG 2K MME	4 0.2 M KBr 25% PEG 2K MME	5 0.2 M KSCN 25% PEG 2K MME	6 0.8 M NaFormate 25% PEG 2K MME
7 0.3 M NaAcetate 15% PEG 4K	8 0.2 M Li_2SO_4 15% PEG 4K	9 0.2 M $MgCl_2$ 15% PEG 4K	10 0. 2M KBr 15% PEG 4K	11 0.2 M KSCN 15% PEG 4K	12 0.8 M NaFormate 15% PEG 4K
13 0.3 M NaAcetate 10% PEG 8K+ 10% PEG 1K	14 0.2 M Li_2SO_4 10% PEG 8K+ 10% PEG 1K	15 0.2 M $MgCl_2$ 10% PEG 8K+ 10% PEG 1K	16 0.2 M KBr 10% PEG 8K+ 10% PEG 1K	17 0.2 M KSCN 10% PEG 8K+ 10% PEG 1K	18 0.8 M NaFormate 10% PEG 8K+ 10% PEG 1K
19 0.3 M NaAcetate 8% PEG 20K+ 8% PEG550 MME	20 0.2 M Li_2SO_4 8% PEG 20K+ 8% PEG550 MME	21 0.2 M $MgCl_2$ 8% PEG 20K+ 8% PEG550 MME	22 0.2 M KBr 8% PEG 20K+ 8% PEG550 MME	23 0.2 M KSCN 8% PEG 20K+ 8% PEG550 MME	24 0.8 M NaFormate 8% PEG 20K+ 8% PEG550 MME

Clear Strategy Screen II (MDL)

1 1. 5M $(NH_4)_2SO_4$	2 0.8M Li_2SO_4	3 2 M NaFormate	4 0.5 M KH_2PO_4	5 25% PEG 2KMME *0.2 M CaAcetate*	6 15% PEG 4K *0.2 M CaAcetate*
7 2. 7M $(NH_4)_2SO_4$	8 1.8M Li_2SO_4	9 4 M NaFormate	10 1.0 M KH_2PO_4	11 10% PEG 8K +10% PEG 1K 0.2 M CaAcetate	12 8% PEG 20K +8% PEG550MME 0.2 M CaAcetate
13 40% v/v MPD	14 40% v/v Butanediol	15 20% PEG 4K 5 mM $CdCl_2$	16 20% PEG 550 MME 0.1 5M KSCN	17 20% PEG 600 0.15 M KSCN	18 20% PEG 1.5K 0.1 5M KSCN
19 35% v/v Isopropanol	20 30% v/v Jeffamine 600M	21 20% PEG 4K 5mM $NiCl_2$	22 18% PEG 3350 0.1 5M KSCN	23 18% PEG 5K MME 0.15 M KSCN	24 15% PEG 6K 0.15 M KSCN

Buffers (1M): Sodium Acetate pH 4.5, pH 5.5; Sodium Cacodylate pH 6.5; Tris pH 7.5, 8.5.

Structure Screen I and II 96 HT (MDL)

tube	salt	buffer	pH	precipitant
A1	0.02 M calcium chloride	0.1 M Na acetate	4.6	30 % v/v MPD
A2	0.2 M ammonium acetate	0.1 M Na acetate	4.6	30 % w/v PEG 4000
A3	0.2 M ammonium sulfate	0.1 M Na acetate	4.6	25 % w/v PEG 4000
A4	None	0.1 M Na acetate	4.6	2.0 M sodium formate
A5	None	0.1 M Na acetate	4.6	2.0 M ammonium sulfate
A6	None	0.1 M Na acetate	4.6	8 % w/v PEG 4000
A7	0.2 M ammonium acetate	0.1 M Na citrate	5.6	30 % w/v PEG 4000
A8	0.2 M ammonium acetate	0.1 M Na citrate	5.6	30 % v/v MPD
A9	None	0.1 M Na citrate	5.6	20 % v/v 2-propanol, 20 % w/v PEG 4000
A10	None	0.1 M Na citrate	5.6	1.0 M ammonium dihydrogen phosphate
A11	0.2 M calcium chloride	0.1 M Na acetate	4.6	20 % v/v 2-propanol
A12	None	0.1 M Na cacodylate	6.5	1.4 M sodium acetate
B1	0.2 M sodium citrate	0.1 M Na cacodylate	6.5	30 % v/v 2-propanol
B2	0.2 M ammonium sulfate	0.1 M Na cacodylate	6.5	30 % w/v PEG 8000
B3	0.2 M magnesium acetate	0.1 M Na cacodylate	6.5	20 % w/v PEG 8000
B4	0.2 M magnesium acetate	0.1 M Na cacodylate	6.5	30 % v/v MPD
B5	None	0.1 M imidazole	6.5	1.0 M sodium acetate
B6	0.2 M sodium acetate	0.1 M Na cacodylate	6.5	30 % w/v PEG 8000
B7	0.2 M zinc acetate	0.1 M Na cacodylate	6.5	18 % w/v PEG 8000
B8	0.2 M calcium acetate	0.1 M Na cacodylate	6.5	18 % w/v PEG 8000
B9	0.2 M sodium citrate	0.1 M Na HEPES	7.5	30 % v/v MPD
B10	0.2 M magnesium chloride	0.1 M Na HEPES	7.5	30 % v/v 2-propanol
B11	0.2 M calcium chloride	0.1 M Na HEPES	7.5	28 % v/v PEG 400
B12	0.2 M magnesium chloride	0.1 M Na HEPES	7.5	30 % v/v PEG 400
C1	0.2 M sodium citrate	0.1 M Na HEPES	7.5	20 % v/v 2-propanol
C2	None	0.1 M Na HEPES	7.5	0.8 M K/Na tartrate
C3	None	0.1 M Na HEPES	7.5	1.5 M lithium sulfate
C4	None	0.1 M Na HEPES	7.5	0.8 M sodium dihydrogen phosphate 0.8 M potassium dihydrogen phosphate
C5	None	0.1 M Na HEPES	7.5	1.4 M tri-sodium citrate
C6	None	0.1 M Na HEPES	7.5	2 % v/v PEG 400, 2.0 M ammonium sulfate
C7	None	0.1 M Na HEPES	7.5	10 % v/v 2-propanol, 20 % w/v PEG 4000
C8	None	0.1 M Tris	8.5	2.0 M ammonium sulfate
C9	0.2 M magnesium chloride	0.1 M Tris	8.5	30 % w/v PEG 4000
C10	0.2 M sodium citrate	0.1 M Tris	8.5	30 % v/v PEG 400
C11	0.2 M lithium sulfate	0.1 M Tris	8.5	30 % w/v PEG 4000
C12	0.2 M ammonium acetate	0.1 M Tris	8.5	30 % v/v 2-propanol
D1	0.2 M sodium acetate	0.1 M Tris	8.5	30 % w/v PEG 4000
D2	None	0.1 M Tris	8.5	8 % w/v PEG 8000
D3	None	0.1 M Tris	8.5	2.0 M ammonium dihydrogen phosphate
D4	None	None	None	0.4 M K/Na tartrate
D5	None	None	None	0.4 M ammonium dihydrogen phosphate
D6	0.2 M ammonium sulfate	None	None	30 % w/v PEG 8000
D7	0.2 M ammonium sulfate	None	None	30 % w/v PEG 4000
D8	None	None	None	2.0 M ammonium sulfate
D9	None	None	None	4.0 M sodium formate
D10	0.05 M potassium dihydrogen phosphate	None	None	20 % w/v PEG 8000
D11	None	None	None	30 % w/v PEG 1500
D12	None	None	None	0.2 M magnesium formate

E1	0.1 M sodium chloride	0.1 M Bicine	9.0	30 % v/v PEG 550 MME
E2	None	0.1 M Bicine	9.0	2.0 M magnesium chloride
E3	2 % v/v dioxane	0.1 M Bicine	9.0	10 % w/v PEG 20,000
E4	0.2 M magnesium chloride	0.1 M Tris	8.5	3.4 M 1,6-hexanediol
E5	None	0.1 M Tris	8.5	25 % v/v tert-Butanol
E6	0.01 M nickel chloride	0.1 M Tris	8.5	1.0 M lithium sulfate
E7	1.5 M ammonium sulfate	0.1 M Tris	8.5	12 % v/v glycerol
E8	0.2 M ammonium phosphate monobasic	0.1 M Tris	8.5	50 % v/v MPD
E9	None	0.1 M Tris	8.5	20 % v/v ethanol
E10	0.01 M nickel chloride	0.1 M Tris	8.5	20 % w/v PEG 2000 MME
E11	0.5 M ammonium sulfate	0.1 M Na HEPES	7.5	30 % v/v MPD
E12	None	0.1 M Na HEPES	7.5	10 % w/v PEG 6000, 5% v/v MPD
F1	None	0.1 M Na HEPES	7.5	20 % v/v Jeffamine M-600
F2	0.1 M sodium chloride	0.1 M Na HEPES	7.5	1.6 M ammonium sulfate
F3	None	0.1 M Na HEPES	7.5	2.0 M ammonium formate
F4	0.05 M cadmium sulfate	0.1 M Na HEPES	7.5	1.0 M sodium acetate
F5	None	0.1 M Na HEPES	7.5	70 % v/v MPD
F6	None	0.1 M Na HEPES	7.5	4.3 M sodium chloride
F7	None	0.1 M Na HEPES	7.5	10 % w/v PEG 8000, 8 % v/v ethylene glycol
F8	None	0.1 M MES	6.5	1.6 M magnesium sulfate
F9	0.1 M sodium dihydrogen phosphate 0.1 M potassium dihydrogen phosphate	0.1 M MES	6.5	2.0 M sodium chloride
F10	None	0.1 M MES	6.5	12 % w/v PEG 20,000
F11	1.6 M ammonium sulfate	0.1 M MES	6.5	10 % v/v dioxane
F12	0.05 M caesium chloride	0.1 M MES	6.5	30 % v/v Jeffamine M-600
G1	0.01 M cobalt chloride	0.1 M MES	6.5	1.8 M ammonium sulfate
G2	*0.2 M ammonium sulfate*	0.1 M MES	6.5	30 % w/v PEG 5000 MME
G3	0.01 M zinc sulfate	*0.1 M MES*	*6.5*	*25 % v/v PEG 550 MME*
G4	None	0.1 M Na HEPES	7.5	*20 % w/v PEG 10,000*
G5	0.2 M K/Na Tartrate	0.1 M Na citrate	5.6	2.0 M ammonium sulfate
G6	0.5 M ammonium sulfate	0.1 M Na citrate	5.6	1.0 M lithium sulfate
G7	0.5 M sodium chloride	0.1 M Na citrate	5.6	4% v/v polyethyleneimine
G8	None	0.1 M Na citrate	5.6	35 % v/v tert-Butanol
G9	0.01 M ferric chloride	0.1 M Na citrate	5.6	10 % v/v Jeffamine M-600
G10	0.01 M manganese chloride	0.1 M Na citrate	5.6	2.5 M 1,6-hexanediol
G11	None	0.1 M Na acetate	4.6	2.0 M sodium chloride
G12	0.2 M sodium chloride	0.1 M Na acetate	4.6	30 % v/v MPD
H1	0.01 M cobalt chloride	0.1 M Na acetate	4.6	1.0 M 1,6-hexanediol
H2	0.1 M cadmium chloride	0.1 M Na acetate	4.6	30 % v/v PEG 400
H3	0.2 M ammonium sulfate	0.1 M Na acetate	4.6	30 % w/v PEG 2000 MME
H4	2.0 M sodium chloride	None	None	10 % w/v PEG 6000
H5	0.01 M CTAB	None	None	0.5 M sodium chloride 0.1 M magnesium chloride
H6	None	None	None	25 % v/v ethylene glycol
H7	None	None	None	35 % v/v dioxane
H8	2.0 M ammonium sulfate	None	None	5 % v/v 2-propanol
H9	None	None	7.0	1.0 M imidazole
H10	None	None	None	10 % w/v PEG 1000, 10% w/v PEG 8000
H11	1.5 M sodium chloride	None	None	10 % v/v ethanol
H12	None	None	6.5	1.6 M sodium citrate

PACT-*premiere* 96 HT (MDL)

tube	salt	buffer	pH	PEG
A1	0.1 M SPG buffer	None	4.0	25 % w/v PEG 1500
A2	0.1 M SPG buffer	None	5.0	25 % w/v PEG 1500
A3	0.1 M SPG buffer	None	6.0	25 % w/v PEG 1500
A4	0.1 M SPG buffer	None	7.0	25 % w/v PEG 1500
A5	0.1 M SPG buffer	None	8.0	25 % w/v PEG 1500
A6	0.1 M SPG buffer	None	9.0	25 % w/v PEG 1500
A7	0.2 M sodium chloride	0.1 M sodium acetate	5.0	20 % w/v PEG 6000
A8	0.2 M ammonium chloride	0.1 M sodium acetate	5.0	20 % w/v PEG 6000
A9	0.2 M lithium chloride	0.1 M sodium acetate	5.0	20 % w/v PEG 6000
A10	0.2 M magnesium chloride	0.1 M sodium acetate	5.0	20 % w/v PEG 6000
A11	0.2 M calcium chloride	0.1 M sodium acetate	5.0	20 % w/v PEG 6000
A12	0.01 M zinc chloride	0.1 M sodium acetate	5.0	20 % w/v PEG 6000
B1	0.1 M MIB buffer	None	4.0	25 % w/v PEG 1500
B2	*0.1 M MIB buffer*	None	5.0	25 % w/v PEG 1500
B3	0.1 M MIB buffer	*None*	*6.0*	25 % w/v PEG 1500
B4	0.1 M MIB buffer	None	7.0	*25 % w/v PEG 1500*
B5	0.1 M MIB buffer	None	8.0	25 % w/v PEG 1500
B6	0.1 M MIB buffer	None	9.0	25 % w/v PEG 1500
B7	0.2 M sodium chloride	0.1 M MES	6.0	20 % w/v PEG 6000
B8	0.2 M ammonium chloride	0.1 M MES	6.0	20 % w/v PEG 6000
B9	0.2 M lithium chloride	0.1 M MES	6.0	20 % w/v PEG 6000
B10	0.2 M magnesium chloride	0.1 M MES	6.0	20 % w/v PEG 6000
B11	0.2 M calcium chloride	0.1 M MES	6.0	20 % w/v PEG 6000
B12	0.01 M zinc chloride	0.1 M MES	6.0	20 % w/v PEG 6000
C1	0.1 M PCTP buffer	None	4.0	25 % w/v PEG 1500
C2	0.1 M PCTP buffer	None	5.0	25 % w/v PEG 1500
C3	0.1 M PCTP buffer	None	6.0	25 % w/v PEG 1500
C4	0.1 M PCTP buffer	None	7.0	25 % w/v PEG 1500
C5	0.1 M PCTP buffer	None	8.0	25 % w/v PEG 1500
C6	0.1 M PCTP buffer	None	9.0	25 % w/v PEG 1500
C7	0.2 M sodium chloride	0.1 M HEPES	7.0	20 % w/v PEG 6000
C8	0.2 M ammonium chloride	0.1 M HEPES	7.0	20 % w/v PEG 6000
C9	0.2 M lithium chloride	0.1 M HEPES	7.0	20 % w/v PEG 6000
C10	0.2 M magnesium chloride	0.1 M HEPES	7.0	20 % w/v PEG 6000
C11	0.2 M calcium chloride	0.1 M HEPES	7.0	20 % w/v PEG 6000
C12	0.01 M zinc chloride	0.1 M HEPES	7.0	20 % w/v PEG 6000
D1	0.1 M MMT buffer	None	4.0	25 % w/v PEG 1500
D2	0.1 M MMT buffer	None	5.0	25 % w/v PEG 1500
D3	0.1 M MMT buffer	None	6.0	25 % w/v PEG 1500
D4	0.1 M MMT buffer	None	7.0	25 % w/v PEG 1500
D5	0.1 M MMT buffer	None	8.0	25 % w/v PEG 1500
D6	0.1 M MMT buffer	None	9.0	25 % w/v PEG 1500
D7	0.2 M sodium chloride	0.1 M Tris	8.0	20 % w/v PEG 6000
D8	0.2 M ammonium chloride	0.1 M Tris	8.0	20 % w/v PEG 6000
D9	0.2 M lithium chloride	0.1 M Tris	8.0	20 % w/v PEG 6000
D10	0.2 M magnesium chloride	0.1 M Tris	8.0	20 % w/v PEG 6000
D11	0.2 M calcium chloride	0.1 M Tris	8.0	20 % w/v PEG 6000
D12	0.002 M zinc chloride	0.1 M Tris	8.0	20 % w/v PEG 6000

E1	0.2 M sodium fluoride	None		20 % w/v PEG 3350
E2	0.2 M sodium bromide	None		20 % w/v PEG 3350
E3	0.2 M sodium iodide	None		20 % w/v PEG 3350
E4	0.2 M potassium thiocyanate	None		20 % w/v PEG 3350
E5	0.2 M sodium nitrate	None		20 % w/v PEG 3350
E6	0.2 M sodium formate	None		20 % w/v PEG 3350
E7	0.2 M sodium acetate	None		20 % w/v PEG 3350
E8	0.2 M sodium sulfate	None		20 % w/v PEG 3350
E9	0.2 M potassium/sodium tartrate	None		20 % w/v PEG 3350
E10	0.02 M sodium/potassium phosphate	None		20 % w/v PEG 3350
E11	0.2 M sodium citrate	None		20 % w/v PEG 3350
E12	0.2 M sodium malonate	None		20 % w/v PEG 3350
F1	0.2 M sodium fluoride	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
F2	0.2 M sodium bromide	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
F3	0.2 M sodium iodide	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
F4	0.2 M potassium thiocyanate	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
F5	0.2 M sodium nitrate	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
F6	0.2 M sodium formate	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
F7	0.2 M sodium acetate	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
F8	0.2 M sodium sulfate	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
F9	0.2 M potassium/sodium tartrate	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
F10	0.02 M sodium/potassium phosphate	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
F11	0.2 M sodium citrate	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
F12	0.2 M sodium malonate	0.1 M Bis Tris propane	6.5	20 % w/v PEG 3350
G1	0.2 M sodium fluoride	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
G2	0.2 M sodium bromide	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
G3	0.2 M sodium iodide	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
G4	0.2 M potassium thiocyanate	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
G5	0.2 M sodium nitrate	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
G6	0.2 M sodium formate	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
G7	0.2 M sodium acetate	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
G8	0.2 M sodium sulfate	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
G9	0.2 M potassium/sodium tartrate	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
G10	0.02 M sodium/potassium phosphate	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
G11	0.2 M sodium citrate	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
G12	0.2 M sodium malonate	0.1 M Bis Tris propane	7.5	20 % w/v PEG 3350
H1	0.2 M sodium fluoride	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350
H2	0.2 M sodium bromide	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350
H3	0.2 M sodium iodide	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350
H4	0.2 M potassium thiocyanate	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350
H5	0.2 M sodium nitrate	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350
H6	0.2 M sodium formate	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350
H7	0.2 M sodium acetate	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350
H8	0.2 M sodium sulfate	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350
H9	0.2 M potassium/sodium tartrate	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350
H10	0.02 M sodium/potassium phosphate	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350
H11	0.2 M sodium citrate	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350
H12	0.2 M sodium malonate	0.1 M Bis Tris propane	8.5	20 % w/v PEG 3350

JCSG-*plus* 96 HT (MDL)

tube	salt	buffer	pH	precipitant
A1	0.2 M lithium sulfate	0.1 M sodium acetate	4.5	50 % v/v PEG 400
A2	None	0.1 M Citrate	5.5	20 % w/v PEG 3K
A3	0.2 M di-ammonium hydrogen citrate	None	-	20 % w/v PEG 3350
A4	0.02 M calcium chloride	0.1 M Sodium acetate	4.6	30 % v/v MPD
A5	0.2 M magnesium formate	None	-	20 % w/v PEG 3350
A6	0.2 M lithium sulfate	0.1 M phosphate/citrate	4.2	20 % w/v PEG 1K
A7	None	0.1 M CHES	9.5	20 % w/v PEG 8K
A8	0.2 M ammonium formate	None	-	20 % w/v PEG 3350
A9	0.2 M ammonium chloride	None	-	20 % w/v PEG 3350
A10	0.2 M potassium formate	None	-	20 % w/v PEG 3350
A11	0.2 M ammonium dihydrogen phosphate	0.1 M Tris	8.5	50 % v/v MPD
A12	0.2 M potassium nitrate	None	-	20 % w/v PEG 3350
B1	None	0.1 M Na Citrate	4.0	0.8 M ammonium sulfate
B2	0.2 M sodium thiocyanate	None	-	20 % w/v PEG 3350
B3	None	0.1 M Bicine	9.0	20 % w/v PEG 6K
B4	None	0.1 M HEPES	7.5	10 % w/v PEG 8K/ 8 % v/v Ethylene glycol
B5	None	0.1 M sodium cacodylate	6.5	40 % v/v MPD/ 5 % w/v PEG 8000
B6	None	0.1 M phosphate/citrate	4.2	40 % v/v Ethanol/ 5 % w/v PEG 1000
B7	None	0.1 M sodium acetate	4.6	8 % w/v PEG 4K
B8	0.2 M magnesium chloride	0.1 M Tris	7.0	10 % w/v PEG 8K
B9	None	0.1 M Na citrate	5.0	20 % w/v PEG 6K
B10	0.2 M magnesium chloride	0.1 M sodium cacodylate	6.5	50 % v/v PEG 200
B11	None	None	6.5	1.6 M tri-sodium citrate
B12	0.2 M tri-potassium citrate	None	-	20 % w/v PEG 3350
C1	0.2 M sodium chloride	0.1 M phosphate/citrate	4.2	20 % w/v PEG 8K
C2	1.0 M lithium chloride	0.1 M Na citrate	4.0	20 % w/v PEG 6K
C3	0.2 M ammonium nitrate	None	-	20 % w/v PEG 3350
C4	None	0.1 M HEPES	7.0	10 % w/v PEG 6K
C5	None	0.1 M Na HEPES	7.5	0.8 M sodium dihydrogen phosphate 0.8 M potassium dihydrogen phosphate
C6	None	0.1 M phosphate/citrate	4.2	40 % v/v PEG 300
C7	0.2 M zinc acetate	0.1 M sodium acetate	4.5	10 % w/v PEG 3K
C8	None	0.1 M Tris	8.5	20 % v/v Ethanol
C9	None	0.1 M Na/K phosphate	6.2	25 % v/v 1,2-propanediol 10 % v/v Glycerol
C10	None	0.1 M Bicine	9.0	10 % w/v PEG 20,000/ 2% v/v Dioxane
C11	None	0.1 M sodium acetate	4.6	2.0 M ammonium sulfate
C12	None	None	-	10 % w/v PEG 1000/ 10 % w/v PEG 8000
D1	None	None	-	24 % w/v PEG 1500/ 20 % v/v Glycerol
D2	0.2 M magnesium chloride	0.1 M Na HEPES	7.5	30 % v/v PEG 400
D3	0.2 M sodium chloride	0.1 M Na/K phosphate	6.2	50 % v/v PEG 200
D4	0.2 M lithium sulfate	0.1 M sodium acetate	4.5	30 % w/v PEG 8K
D5	None	0.1 M HEPES	7.5	70 % v/v MPD
D6	0.2 M magnesium chloride	0.1 M Tris	8.5	20 % w/v PEG 8K
D7	*0.2 M lithium sulfate*	0.1 M Tris	8.5	40 % v/v PEG 400
D8	None	*0.1 M Tris*	8.0	40 % v/v MPD
D9	0.17 M ammonium sulfate	None	-	*25.5 % w/v PEG 4K/ 15 % v/v Glycerol*
D10	0.2 M calcium acetate	0.1 M Sodium cacodylate	6.5	40 % v/v PEG 300
D11	0.14 M calcium chloride	0.07 M Sodium acetate	4.6	14 % v/v 2-propanol/ 30 % v/v Glycerol
D12	0.04 M potassium dihydrogen phosphate	None	-	16 % w/v PEG 8K/ 20 % v/v Glycerol

E1	None	0.1 M sodium cacodylate	6.5	1.0 M tri-sodium citrate
E2	0.2 M sodium chloride	0.1 M sodium cacodylate	6.5	2.0 M ammonium sulfate
E3	0.2 M sodium chloride	0.1 M HEPES	7.5	10 % v/v 2-propanol
E4	0.2 M lithium sulfate	0.1 M Tris	8.5	1.26 M ammonium sulfate
E5	None	0.1 M CAPS	10.5	40 % v/v MPD
E6	0.2 M zinc acetate	0.1 M imidazole	8.0	20 % w/v PEG 3K
E7	0.2 M zinc acetate	0.1 M sodium cacodylate	6.5	10 % v/v 2-propanol
E8	None	0.1 M sodium acetate	4.5	1.0 M di-ammonium hydrogen phosphate
E9	None	0.1 M MES	6.5	1.6 M magnesium sulfate
E10	None	0.1 M Bicine	9.0	10 % w/v PEG 6K
E11	0.16 M calcium acetate	0.08 M sodium cacodylate	6.5	14.4 % w/v PEG 8K/ 20 % v/v glycerol
E12	None	0.1 M imidazole	8.0	10 % w/v PEG 8K
F1	0.05 M caesium chloride	0.1 M MES	6.5	30 % v/v Jeffamine M-600
F2	None	0.1 M Na Citrate	5.0	3.2 M ammonium sulfate
F3	None	0.1 M Tris	8.0	20 % v/v MPD
F4	None	0.1 M HEPES	7.5	20 % v/v Jeffamine M-600
F5	0.2 M magnesium chloride	0.1 M Tris	8.5	50 % v/v ethylene glycol
F6	None	0.1 M Bicine	9.0	10 % v/v MPD
F7	None	None	7.0	0.8 M succinic acid
F8	None	None	7.0	2.1 M DL-malic acid
F9	None	None	7.0	2.4 M sodium malonate
F10	1.1 M sodium malonate	0.1 M HEPES	7.0	0.5 % v/v Jeffamine ED-2001
F11	1.0 M succinic acid	0.1 M HEPES	7.0	1 % w/v PEG 2000 MME
F12	None	0.1 M HEPES	7.0	30 % v/v Jeffamine M-600
G1	None	0.1 M HEPES	7.0	30 % v/v Jeffamine ED-2001
G2	0.02 M magnesium chloride	0.1 M HEPES	7.5	22 % w/v polyacrylic acid 5100 sodium salt
G3	0.01 M cobalt chloride	0.1 M Tris	8.5	20 % w/v polyvinylpyrrolidone K15
G4	0.2 M tri-methylamine N-oxide	0.1 M Tris	8.5	20 % w/v PEG 2K MME
G5	0.005 M cobalt chloride 0.005 M cadmium chloride 0.005 M magnesium chloride 0.005 M nickel chloride	0.1 M HEPES	7.5	12 % w/v PEG 3350
G6	0.2 M sodium malonate	None	7.0	20 % w/v PEG 3350
G7	0.1 M succinic acid	None	7.0	15 % w/v PEG 3350
G8	0.15 M DL- malic acid	None	7.0	20 % w/v PEG 3350
G9	0.1 M potassium thiocyanate	None	None	30 % w/v PEG 2K MME
G10	0.15 M potassium bromide	None	None	30 % w/v PEG 2K MME
G11	None	0.1 M Bis Tris	5.5	2.0 M ammonium sulfate
G12	None	0.1 M Bis Tris	5.5	3.0 M sodium chloride
H1	None	0.1 M Bis Tris	5.5	0.3 M magnesium formate
H2	1.0 M ammonium sulfate	0.1 M Bis Tris	5.5	1 % w/v PEG 3350
H3	None	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
H4	0.2 M calcium chloride	0.1 M Bis Tris	5.5	45 % v/v MPD
H5	0.2 M ammonium acetate	0.1 M Bis Tris	5.5	45 % v/v MPD
H6	0.1 M ammonium acetate	0.1 M Bis Tris	5.5	17 % w/v PEG 10K
H7	0.2 M ammonium sulfate	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
H8	0.2 M sodium chloride	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
H9	0.2 M lithium sulfate	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
H10	0.2 M ammonium acetate	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
H11	0.2 M magnesium chloride	0.1 M Bis Tris	5.5	25 % w/v PEG 3350
H12	0.2 M ammonium acetate	0.1 M HEPES	7.5	45 % v/v MPD

Bibliography

1. Fleming, A., *On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzæ.* Br J Exp Pathol 1929. **31**: p. 226-36.
2. Demain, A.L., *Small bugs, big business: the economic power of the microbe.* Biotechnol Adv, 2000. **18**(6): p. 499-514.
3. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review.* Natural Product Reports, 2001. **18**(4): p. 380-416.
4. Hopwood, D.A., *Molecular-Genetics of Polyketides and Its Comparison to Fatty- Acid Biosynthesis.* Annual Review of Genetics, 1990. **24**: p. 37-66.
5. Carreras, C.W., R. Pieper, and C. Khosla, *The chemistry and biology of fatty acid, polyketide, and nonribosomal peptide biosynthesis*, in *Bioorganic Chemistry Deoxysugars, Polyketides and Related Classes: Synthesis, Biosynthesis, Enzymes*. 1997. p. 85-126.
6. Wakil, S.J., *Fatty acid synthase, a proficient multifunctional enzyme.* Biochemistry, 1989. **28**(11): p. 4523-30.
7. Mann, J., *Secondary Metabolism.* Clarendon Press: Oxford, 1992. **2nd edition.**
8. Smith, S., *The animal fatty acid synthase: one gene, one polypeptide, seven enzymes.* Faseb J, 1994. **8**(15): p. 1248-59.
9. Smith, S., A. Witkowski, and A.K. Joshi, *Structural and functional organization of the animal fatty acid synthase.* Prog Lipid Res, 2003. **42**(4): p. 289-317.
10. Chirala, S.S. and S.J. Wakil, *Structure and function of animal fatty acid synthase.* Lipids, 2004. **39**(11): p. 1045-53.
11. Witkowski, A., A.K. Joshi, V.S. Rangan, A.M. Falick, H.E. Witkowska, and S. Smith, *Dibromopropanone cross-linking of the phosphopantetheine and active-site cysteine thiols of the animal fatty acid synthase can occur both inter- and intrasubunit. Reevaluation of the side-by-side, antiparallel subunit model.* J Biol Chem, 1999. **274**(17): p. 11557-63.
12. Rangan, V.S., A.K. Joshi, and S. Smith, *Mapping the functional topology of the animal fatty acid synthase by mutant complementation in vitro.* Biochemistry, 2001. **40**(36): p. 10792-9.
13. Joshi, A.K., V.S. Rangan, A. Witkowski, and S. Smith, *Engineering of an active animal fatty acid synthase dimer with only one competent subunit.* Chem Biol, 2003. **10**(2): p. 169-73.
14. Maier, T., S. Jenni, and N. Ban, *Architecture of mammalian fatty acid synthase at 4.5 A resolution.* Science, 2006. **311**(5765): p. 1258-62.
15. Asturias, F.J., J.Z. Chadick, I.K. Cheung, H. Stark, A. Witkowski, A.K. Joshi, and S. Smith, *Structure and molecular organization of mammalian fatty acid synthase.* Nat Struct Mol Biol, 2005. **12**(3): p. 225-32.
16. Jenni, S., M. Leibundgut, T. Maier, and N. Ban, *Architecture of a fungal fatty acid synthase at 5 A resolution.* Science, 2006. **311**(5765): p. 1263-7.
17. Johansson, P., B. Wiltschi, P. Kumari, B. Kessler, C. Vonrhein, J. Vonck, D. Oesterhelt, and M. Grininger, *Inhibition of the fungal fatty acid synthase type I multienzyme complex.* Proc Natl Acad Sci U S A, 2008. **105**(35): p. 12803-8.

18. Leibundgut, M., S. Jenni, C. Frick, and N. Ban, *Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase.* Science, 2007. **316**(5822): p. 288-90.
19. Lomakin, I.B., Y. Xiong, and T.A. Steitz, *The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together.* Cell, 2007. **129**(2): p. 319-32.
20. Jenni, S., M. Leibundgut, D. Boehringer, C. Frick, B. Mikolasek, and N. Ban, *Structure of fungal fatty acid synthase and implications for iterative substrate shuttling.* Science, 2007. **316**(5822): p. 254-61.
21. Serre, L., E.C. Verbree, Z. Dauter, A.R. Stuitje, and Z.S. Derewenda, *The Escherichia coli malonyl-CoA:acyl carrier protein transacylase at 1.5-A resolution. Crystal structure of a fatty acid synthase component.* J Biol Chem, 1995. **270**(22): p. 12961-4.
22. Serre, L., L. Swenson, R. Green, Y. Wei, Verwoert, II, E.C. Verbree, A.R. Stuitje, and Z.S. Derewenda, *Crystallization of the malonyl coenzyme A-acyl carrier protein transacylase from Escherichia coli.* J Mol Biol, 1994. **242**(1): p. 99-102.
23. Jackowski, S. and C.O. Rock, *Acetoacetyl-acyl carrier protein synthase, a potential regulator of fatty acid biosynthesis in bacteria.* J Biol Chem, 1987. **262**(16): p. 7927-31.
24. Jornvall, H., B. Persson, M. Krook, S. Atrian, R. Gonzalez-Duarte, J. Jeffery, and D. Ghosh, *Short-chain dehydrogenases/reductases (SDR).* Biochemistry, 1995. **34**(18): p. 6003-13.
25. Zhang, Y.M., B.N. Wu, J. Zheng, and C.O. Rock, *Key residues responsible for acyl carrier protein and beta- ketoacyl-acyl carrier protein reductase (FabG) interaction.* Journal of Biological Chemistry, 2003. **278**(52): p. 52935-52943.
26. Heath, R.J. and C.O. Rock, *Roles of the FabA and FabZ beta-hydroxyacyl-acyl carrier protein dehydratases in Escherichia coli fatty acid biosynthesis.* J Biol Chem, 1996. **271**(44): p. 27795-801.
27. Heath, R.J., N. Su, C.K. Murphy, and C.O. Rock, *The enoyl-[acyl-carrier-protein] reductases FabI and FabL from Bacillus subtilis.* J Biol Chem, 2000. **275**(51): p. 40128-33.
28. Heath, R.J. and C.O. Rock, *Regulation of malonyl-CoA metabolism by acyl-acyl carrier protein and beta-ketoacyl-acyl carrier protein synthases in Escherichia coli.* J Biol Chem, 1995. **270**(26): p. 15531-8.
29. Hertweck, C., A. Luzhetskyy, Y. Rebets, and A. Bechthold, *Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork.* Nat Prod Rep, 2007. **24**(1): p. 162-90.
30. Funa, N., Y. Ohnishi, I. Fujii, M. Shibuya, Y. Ebizuka, and S. Horinouchi, *A new pathway for polyketide synthesis in microorganisms.* Nature, 1999. **400**(6747): p. 897-9.
31. Jez, J.M., M.B. Austin, J. Ferrer, M.E. Bowman, J. Schroder, and J.P. Noel, *Structural control of polyketide formation in plant-specific polyketide synthases.* Chem Biol, 2000. **7**(12): p. 919-30.
32. Jez, J.M., M.E. Bowman, R.A. Dixon, and J.P. Noel, *Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase.* Nat Struct Biol, 2000. **7**(9): p. 786-91.

33. Jez, J.M. and J.P. Noel, *Mechanism of chalcone synthase. pKa of the catalytic cysteine and the role of the conserved histidine in a plant polyketide synthase.* J Biol Chem, 2000. **275**(50): p. 39640-6.
34. Hutchinson, C.R., J. Kennedy, C. Park, S. Kendrew, K. Auclair, and J. Vederas, *Aspects of the biosynthesis of non-aromatic fungal polyketides by iterative polyketide synthases.* Antonie Van Leeuwenhoek, 2000. **78**(3-4): p. 287-95.
35. Jordan, P.M. and J.B. Spencer, *The biosynthesis of tetraketides: enzymology, mechanism and molecular programming.* Biochem Soc Trans, 1993. **21**(1): p. 222-8.
36. Hendrickson, L., C.R. Davis, C. Roach, D.K. Nguyen, T. Aldrich, P.C. McAda, and C.D. Reeves, *Lovastatin biosynthesis in Aspergillus terreus: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene.* Chem Biol, 1999. **6**(7): p. 429-39.
37. Kennedy, J., K. Auclair, S.G. Kendrew, C. Park, J.C. Vederas, and C.R. Hutchinson, *Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis.* Science, 1999. **284**(5418): p. 1368-72.
38. Dawson, M.J., J.E. Farthing, P.S. Marshall, R.F. Middleton, M.J. O'Neill, A. Shuttleworth, C. Stylli, R.M. Tait, P.M. Taylor, H.G. Wildman, *The squalestatins, novel inhibitors of squalene synthase produced by a species of Phoma. I. Taxonomy, fermentation, isolation, physico-chemical properties and biological activity.* J Antibiot (Tokyo), 1992. **45**(5): p. 639-47.
39. Cox, R.J., Glod, F., Hurley, D., Lazarus, C.M., Nicholson, T.P., Rudd, B.A., Simpson, T.J., Wilkinson, B. and Zhang, Y., *Rapid cloning and expression of a fungal polyketide synthase gene involved in suqlestatin biosynthesis.* Chem Cummun, 2004: p. 2260-2261.
40. Staunton, J. and B. Wilkinson, *Biosynthesis of erythromycin and rapamycin.* Chemical Reviews, 1997. **97**(7): p. 2611-2629.
41. Donadio, S., J.B. McAlpine, P.J. Sheldon, M. Jackson, and L. Katz, *An erythromycin analog produced by reprogramming of polyketide synthesis.* Proc Natl Acad Sci U S A, 1993. **90**(15): p. 7119-23.
42. Donadio, S. and M.J. Staver, *IS1136, an insertion element in the erythromycin gene cluster of Saccharopolyspora erythraea.* Gene, 1993. **126**(1): p. 147-51.
43. Bevitt, D.J., J. Cortes, S.F. Haydock, and P.F. Leadlay, *6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme.* Eur J Biochem, 1992. **204**(1): p. 39-49.
44. Donadio, S., M.J. Staver, J.B. McAlpine, S.J. Swanson, and L. Katz, *Modular organization of genes required for complex polyketide biosynthesis.* Science, 1991. **252**(5006): p. 675-9.
45. Donadio, S. and L. Katz, *Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in Saccharopolyspora erythraea.* Gene, 1992. **111**(1): p. 51-60.

46. Staunton, J., P. Caffrey, J.F. Aparicio, G.A. Roberts, S.S. Bethell, and P.F. Leadlay, *Evidence for a double-helical structure for modular polyketide synthases.* Nat Struct Biol, 1996. **3**(2): p. 188-92.
47. Carreras, C.W. and C. Khosla, *Purification and in vitro reconstitution of the essential protein components of an aromatic polyketide synthase.* Biochemistry, 1998. **37**(8): p. 2084-8.
48. Carreras, C.W., Pieper, R. and Khosla, C., *Efficient Synthesis of Aromatic Polyketides in Vitro by the Actinorhodin Polyketide Synthase.* J. Am. Chem. Soc, 1996. **118**: p. 5158-5159.
49. Fernandez-Moreno, M.A., E. Martinez, L. Boto, D.A. Hopwood, and F. Malpartida, *Nucleotide sequence and deduced functions of a set of cotranscribed genes of Streptomyces coelicolor A3(2) including the polyketide synthase for the antibiotic actinorhodin.* J Biol Chem, 1992. **267**(27): p. 19278-90.
50. McDaniel, R., S. Ebertkhosla, H. Fu, D.A. Hopwood, and C. Khosla, *Engineered Biosynthesis of Novel Polyketides - Influence of a Downstream Enzyme on the Catalytic Specificity of a Minimal Aromatic Polyketide Synthase.* Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(24): p. 11542-11546.
51. Kramer, P.J., R.J.X. Zawada, R. McDaniel, C.R. Hutchinson, D.A. Hopwood, and C. Khosla, *Rational design and engineered biosynthesis of a novel 18- carbon aromatic polyketide.* Journal of the American Chemical Society, 1997. **119**(4): p. 635-639.
52. McDaniel, R., S. Ebertkhosla, D.A. Hopwood, and C. Khosla, *Engineered Biosynthesis of Novel Polyketides - Act(Vii) and Act(Iv) Genes Encode Aromatase and Cyclase Enzymes, Respectively.* Journal of the American Chemical Society, 1994. **116**(24): p. 10855-10859.
53. Fu, H., D.A. Hopwood, and C. Khosla, *Engineered biosynthesis of novel polyketides: evidence for temporal, but not regiospecific, control of cyclization of an aromatic polyketide precursor.* Chem Biol, 1994. **1**(4): p. 205-10.
54. Bisang, C., P.F. Long, J. Cortes, J. Westcott, J. Crosby, A.L. Matharu, R.J. Cox, T.J. Simpson, J. Staunton, and P.F. Leadlay, *A chain initiation factor common to both modular and aromatic polyketide synthases.* Nature, 1999. **401**(6752): p. 502-5.
55. Shen, B., R.G. Summers, H. Gramajo, M.J. Bibb, and C.R. Hutchinson, *Purification and characterization of the acyl carrier protein of the Streptomyces glaucescens tetracenomycin C polyketide synthase.* J Bacteriol, 1992. **174**(11): p. 3818-21.
56. Hopwood, D.A., *Genetic contributions to understanding polyketide synthases.* Chemical Reviews, 1997. **97**(7): p. 2465-2497.
57. Wakil, S.J., J.K. Stoops, and V.C. Joshi, *Fatty acid synthesis and its regulation.* Annu Rev Biochem, 1983. **52**: p. 537-79.
58. Lynen, F., *On the structure of fatty acid synthetase of yeast.* Eur J Biochem, 1980. **112**(3): p. 431-42.
59. Kleinkauf, H. and H. Von Dohren, *A nonribosomal system of peptide biosynthesis.* Eur J Biochem, 1996. **236**(2): p. 335-51.

60. Carreras, C.W., A.M. Gehring, C.T. Walsh, and C. Khosla, *Utilization of enzymatically phosphopantetheinylated acyl carrier proteins and acetyl-acyl carrier proteins by the actinorhodin polyketide synthase.* Biochemistry, 1997. **36**(39): p. 11757-61.
61. Lambalot, R.H., A.M. Gehring, R.S. Flugel, P. Zuber, M. LaCelle, M.A. Marahiel, R. Reid, C. Khosla, and C.T. Walsh, *A new enzyme superfamily - the phosphopantetheinyl transferases.* Chem Biol, 1996. **3**(11): p. 923-36.
62. Kim, Y. and J.H. Prestegard, *Refinement of the NMR structures for acyl carrier protein with scalar coupling data.* Proteins, 1990. **8**(4): p. 377-85.
63. Xu, G.Y., A. Tam, L. Lin, J. Hixon, C.C. Fritz, and R. Powers, *Solution structure of B. subtilis acyl carrier protein.* Structure, 2001. **9**(4): p. 277-87.
64. Crump, M.P., J. Crosby, C.E. Dempsey, J.A. Parkinson, M. Murray, D.A. Hopwood, and T.J. Simpson, *Solution structure of the actinorhodin polyketide synthase acyl carrier protein from Streptomyces coelicolor A3(2).* Biochemistry, 1997. **36**(20): p. 6000-6008.
65. Reed, M.A., M. Schweizer, A.E. Szafranska, C. Arthur, T.P. Nicholson, R.J. Cox, J. Crosby, M.P. Crump, and T.J. Simpson, *The type I rat fatty acid synthase ACP shows structural homology and analogous biochemical properties to type II ACPs.* Org Biomol Chem, 2003. **1**(3): p. 463-71.
66. Bunkoczi, G., S. Pasta, A. Joshi, X. Wu, K.L. Kavanagh, S. Smith, and U. Oppermann, *Mechanism and substrate recognition of human holo ACP synthase.* Chem Biol, 2007. **14**(11): p. 1243-53.
67. Findlow, S.C., C. Winsor, T.J. Simpson, J. Crosby, and M.P. Crump, *Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from Streptomyces rimosus.* Biochemistry, 2003. **42**(28): p. 8423-8433.
68. Li, Q., C. Khosla, J.D. Puglisi, and C.W. Liu, *Solution structure and backbone dynamics of the holo form of the frenolicin acyl carrier protein.* Biochemistry, 2003. **42**(16): p. 4648-4657.
69. Crump, M.P., J. Crosby, C.E. Dempsey, M. Murray, D.A. Hopwood, and T.J. Simpson, *Conserved secondary structure in the actinorhodin polyketide synthase acyl carrier protein from Streptomyces coelicolor A3(2) and the fatty acid synthase acyl carrier protein from Escherichia coli.* FEBS Lett, 1996. **391**(3): p. 302-6.
70. Hitchman, T.S., J. Crosby, K.J. Byrom, R.J. Cox, and T.J. Simpson, *Catalytic self-acylation of type II polyketide synthase acyl carrier proteins.* Chem Biol, 1998. **5**(1): p. 35-47.
71. Bao, W., E. Wendt-Pienkowski, and C.R. Hutchinson, *Reconstitution of the iterative type II polyketide synthase for tetracenomycin F2 biosynthesis.* Biochemistry, 1998. **37**(22): p. 8132-8.
72. Arthur, C., R.J. Cox, J. Crosby, M.M. Rahman, T.J. Simpson, F. Soulas, R. Spogli, A.E. Szafranska, J. Westcott, and C.J. Winfield, *Synthesis and characterisation of acyl carrier protein bound polyketide analogues.* Chembiochem, 2002. **3**(2-3): p. 253-7.
73. Elovson, J. and P.R. Vagelos, *Acyl carrier protein. X. Acyl carrier protein synthetase.* J Biol Chem, 1968. **243**(13): p. 3603-11.

74. Cox, R.J., J. Crosby, O. Daltrop, F. Glod, M.E. Jarzabek, T.P. Nicholson, M. Reed, T.J. Simpson, L.H. Smith, F. Soulas, A.E. Szafranska, and J. Westcott, *Streptomyces coelicolor phosphopantetheinyl transferase: a promiscuous activator of polyketide and fatty acid synthase acyl carrier proteins.* Journal of the Chemical Society-Perkin Transactions 1, 2002(14): p. 1644-1649.
75. Gehring, A.M., R.H. Lambalot, K.W. Vogel, D.G. Drueckhammer, and C.T. Walsh, *Ability of Streptomyces spp. acyl carrier proteins and coenzyme A analogs to serve as substrates in vitro for E. coli holo-ACP synthase.* Chem Biol, 1997. 4(1): p. 17-24.
76. Quadri, L.E., P.H. Weinreb, M. Lei, M.M. Nakano, P. Zuber, and C.T. Walsh, *Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases.* Biochemistry, 1998. **37**(6): p. 1585-95.
77. Kealey, J.T., L. Liu, D.V. Santi, M.C. Betlach, and P.J. Barr, *Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts.* Proc Natl Acad Sci U S A, 1998. **95**(2): p. 505-9.
78. Reuter, K., M.R. Mofid, M.A. Marahiel, and R. Ficner, *Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily.* Embo J, 1999. **18**(23): p. 6823-31.
79. Fichtlscherer, F., C. Wellein, M. Mittag, and E. Schweizer, *A novel function of yeast fatty acid synthase. Subunit alpha is capable of self-pantetheinylation.* Eur J Biochem, 2000. **267**(9): p. 2666-71.
80. Flugel, R.S., Y. Hwangbo, R.H. Lambalot, J.E. Cronan, Jr., and C.T. Walsh, *Holo-(acyl carrier protein) synthase and phosphopantetheinyl transfer in Escherichia coli.* J Biol Chem, 2000. **275**(2): p. 959-68.
81. Lam, H.M., E. Tancula, W.B. Dempsey, and M.E. Winkler, *Suppression of insertions in the complex pdxJ operon of Escherichia coli K-12 by lon and other mutations.* J Bacteriol, 1992. **174**(5): p. 1554-67.
82. Takiff, H.E., T. Baker, T. Copeland, S.M. Chen, and D.L. Court, *Locating essential Escherichia coli genes by using mini-Tn10 transposons: the pdxJ operon.* J Bacteriol, 1992. **174**(5): p. 1544-53.
83. Crosby, J., *D.H. Sherman, M.J. Bibb, W.P. Revill, D.A. Hopwood, and T.J.* Simpson, *Polyketide synthase acyl carrier proteins from Streptomyces: expression in Escherichia coli, purification and partial characterisation.* Biochim Biophys Acta, 1995. **1251**(1): p. 32-42.
84. Tropf, S., W.P. Revill, M.J. Bibb, D.A. Hopwood, and M. Schweizer, *Heterologously expressed acyl carrier protein domain of rat fatty acid synthase functions in Escherichia coli fatty acid synthase and Streptomyces coelicolor polyketide synthase systems.* Chem Biol, 1998. **5**(3): p. 135-46.
85. Cox, R.J., T.S. Hitchman, K.J. Byrom, I.S. Findlow, J.A. Tanner, J. Crosby, and T.J. Simpson, *Post-translational modification of heterologously expressed Streptomyces type II polyketide synthase acyl carrier proteins.* FEBS Lett, 1997. **405**(3): p. 267-72.

86. Parris, K.D., L. Lin, A. Tam, R. Mathew, J. Hixon, M. Stahl, C.C. Fritz, J. Seehra, and W.S. Somers, *Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites.* Structure with Folding & Design, 2000. **8**(8): p. 883-895.
87. Chirgadze, N.Y., Briggs, S.L., McAllister, K.A., Fischl, A.S., Zhao, G., *Crystal structure of Streptococcus pneumoniae acyl carrier protein synthase: an essential enzyme in bacterial fatty acid biosynthesis.* Embo J, 2000. **19**: p. 5281-7.
88. Challis, G.L.a.C., K. F., *Incorporation of [U-13C]glycerol defines plausible early steps for the biosynthesis of methylenomycin A in Streptomyces coelicolor A3(2).* Chem. Commun., 2001: p. 935.
89. Cerdeno, A.M., M.J. Bibb, and G.L. Challis, *Analysis of the prodiginine biosynthesis gene cluster of Streptomyces coelicolor A3(2): new mechanisms for chain initiation and termination in modular multienzymes.* Chem Biol, 2001. **8**(8): p. 817-29.
90. Lakey, J.H., E.J. Lea, B.A. Rudd, H.M. Wright, and D.A. Hopwood, *A new channel-forming antibiotic from Streptomyces coelicolor A3(2) which requires calcium for its activity.* J Gen Microbiol, 1983. **129**(12): p. 3565-73.
91. Hopwood, D.A. and H.M. Wright, *CDA is a new chromosomally-determined antibiotic from Streptomyces coelicolor A3(2).* J Gen Microbiol, 1983. **129**(12): p. 3575-9.
92. Chong, P.P., S.M. Podmore, H.M. Kieser, M. Redenbach, K. Turgay, M. Marahiel, D.A. Hopwood, and C.P. Smith, *Physical identification of a chromosomal locus encoding biosynthetic genes for the lipopeptide calcium-dependent antibiotic (CDA) of Streptomyces coelicolor A3(2).* Microbiology, 1998. **144 (Pt 1)**: p. 193-9.
93. Keatinge-Clay, A.T., D.A. Maltby, K.F. Medzihradszky, C. Khosla, and R.M. Stroud, *An antibiotic factory caught in action.* Nat Struct Mol Biol, 2004. **11**(9): p. 888-93.
94. Witkowski, A., A.K. Joshi, Y. Lindqvist, and S. Smith, *Conversion of a beta-ketoacyl synthase to a malonyl decarboxylase by replacement of the active-site cysteine with glutamine.* Biochemistry, 1999. **38**(36): p. 11643-50.
95. McDaniel, R., S. Ebertkhosla, D.A. Hopwood, and C. Khosla, *Engineered Biosynthesis of Novel Polyketides.* Science, 1993. **262**(5139): p. 1546-1550.
96. Burson, K.K.a.K., C., *Dissecting the Chain Length Specificity in Bacterial Aromatic Polyketide Synthases using Chimeric Genes.* Tetrahedron, 2000. **56**(48): p. 9401-9408.
97. Rawlings, B.J., *Biosynthesis of polyketides (other than actinomycete macrolides).* Nat Prod Rep, 1999. **16**(4): p. 425-84.
98. Fu, H., R. McDaniel, D.A. Hopwood, and C. Khosla, *Engineered Biosynthesis of Novel Polyketides - Stereochemical Course of 2 Reactions Catalyzed by a Polyketide Synthase.* Biochemistry, 1994. **33**(31): p. 9321-9326.

99. Fu, H., S. Ebertkhosla, D.A. Hopwood, and C. Khosla, *Engineered Biosynthesis of Novel Polyketides - Dissection of the Catalytic Specificity of the Act Ketoreductase.* Journal of the American Chemical Society, 1994. **116**(10): p. 4166-4170.
100. Zhang, H.-I., X.-g. He, A. Adefarati, J. Gallucci, S.P. Cole, J.M. Beale, P.J. Keller, C.-j. Chang, and H.G. Floss, *Mutactin, a novel polyketide from Streptomyces coelicolor. Structure and biosynthetic relationship to Actinorhodin.* J. Org. Chem., 1990. **55**: p. 1682-1684.
101. Kunnari, T., K. Ylihonko, A. Hautala, K.D. Klika, P. Mantsala, and J. Hakala, *Incorrectly folded aromatic polyketides from polyketide reductase deficient mutants.* Bioorganic & Medicinal Chemistry Letters, 1999. **9**(18): p. 2639-2642.
102. Zawada, R.J.X. and C. Khosla, *Heterologous expression, purification, reconstitution and kinetic analysis of an extended type II polyketide synthase.* Chemistry & Biology, 1999. **6**(9): p. 607-615.
103. Hadfield, A.T., C. Limpkin, W. Teartasin, T.J. Simpson, J. Crosby, and M.P. Crump, *The crystal structure of the actIII actinorhodin polyketide reductase: Proposed mechanism for ACP and polyketide binding.* Structure, 2004. **12**(10): p. 1865-1875.
104. Kallberg, Y., U. Oppermann, H. Jornvall, and B. Persson, *Short-chain dehydrogenase/reductase (SDR) relationships: A large family with eight clusters common to human, animal, and plant genomes.* Protein Science, 2002. **11**(3): p. 636-641.
105. Oppermann, U.C.T., C. Filling, and H. Jornvall, *Forms and functions of human SDR enzymes.* Chemico-Biological Interactions, 2001. **130**(1-3): p. 699-705.
106. Oppermann, U., C. Filling, M. Hult, N. Shafqat, X.Q. Wu, M. Lindh, J. Shafqat, E. Nordling, Y. Kallberg, B. Persson, and H. Jornvall, *Short-chain dehydrogenases/reductases (SDR): the 2002 update.* Chemico-Biological Interactions, 2003. **143**: p. 247-253.
107. Cohen-Gonsaud, M., S. Ducasse, F. Hoh, D. Zerbib, G. Labesse, and A. Quemard, *Crystal structure of MabA from Mycobacterium tuberculosis, a reductase involved in long-chain fatty acid biosynthesis.* Journal of Molecular Biology, 2002. **320**(2): p. 249-261.
108. Fisher, M., J.T.M. Kroon, W. Martindale, A.R. Stuitje, A.R. Slabas, and J.B. Rafferty, *The X-ray structure of Brassica napus beta-keto acyl carrier protein reductase and its implications for substrate binding and catalysis.* Structure with Folding & Design, 2000. **8**(4): p. 339-347.
109. Price, A.C., Y.M. Zhang, C.O. Rock, and S.W. White, *Structure of beta-ketoacyl- acyl carrier protein reductase from Escherichia coli: Negative cooperativity and its structural basis.* Biochemistry, 2001. **40**(43): p. 12772-12781.
110. Price, A.C., Y.M. Zhang, C.O. Rock, and S.W. White, *Cofactor-induced conformational rearrangements establish a catalytically competent active site and a proton relay conduit in FabG.* Structure, 2004. **12**: p. 417-428.
111. Ghosh, D., Z. Wawrzak, C.M. Weeks, W.L. Duax, and M. Erman, *The Refined 3-Dimensional Structure of 3-Alpha,20-Beta- Hydroxysteroid Dehydrogenase and Possible Roles of the Residues Conserved in Short-Chain Dehydrogenases.* Structure, 1994. **2**(7): p. 629-640.

112. Bibb, M.J., D.H. Sherman, S. Omura, and D.A. Hopwood, *Cloning, sequencing and deduced functions of a cluster of Streptomyces genes probably encoding biosynthesis of the polyketide antibiotic frenolicin.* Gene, 1994. **142**(1): p. 31-9.
113. Zawada, R.J. and C. Khosla, *Domain analysis of the molecular recognition features of aromatic polyketide synthase subunits.* J Biol Chem, 1997. **272**(26): p. 16184-8.
114. Thompson, T.B., K. Katayama, K. Watanabe, C.R. Hutchinson, and I. Rayment, *Structural and functional analysis of tetracenomycin F2 cyclase from Streptomyces glaucescens. A type II polyketide cyclase.* J Biol Chem, 2004. **279**(36): p. 37956-63.
115. Mayer, A., T. Taguchi, A. Linnenbrink, C. Hofmann, A. Luzhetskyy, and A. Bechthold, *LanV, a bifunctional enzyme: aromatase and ketoreductase during landomycin A biosynthesis.* Chembiochem, 2005. **6**(12): p. 2312-5.
116. Ames, B.D., T.P. Korman, W. Zhang, P. Smith, T. Vu, Y. Tang, and S.C. Tsai, *Crystal structure and functional analysis of tetracenomycin ARO/CYC: implications for cyclization specificity of aromatic polyketides.* Proc Natl Acad Sci U S A, 2008. **105**(14): p. 5349-54.
117. Lomovskaya, N., S.L. Otten, Y. Doi-Katayama, L. Fonstein, X.C. Liu, T. Takatsu, A. Inventi-Solari, S. Filippini, F. Torti, A.L. Colombo, and C.R. Hutchinson, *Doxorubicin overproduction in Streptomyces peucetius: cloning and characterization of the dnrU ketoreductase and dnrV genes and the doxA cytochrome P-450 hydroxylase gene.* J Bacteriol, 1999. **181**(1): p. 305-18.
118. Grimm, A., K. Madduri, A. Ali, and C.R. Hutchinson, *Characterization of the Streptomyces peucetius ATCC 29050 genes encoding doxorubicin polyketide synthase.* Gene, 1994. **151**(1-2): p. 1-10.
119. Dickens, M.L. and W.R. Strohl, *Isolation and characterization of a gene from Streptomyces sp. strain C5 that confers the ability to convert daunomycin to doxorubicin on Streptomyces lividans TK24.* J Bacteriol, 1996. **178**(11): p. 3389-95.
120. Walczak, R.J., M.L. Dickens, N.D. Priestley, and W.R. Strohl, *Purification, properties, and characterization of recombinant Streptomyces sp. strain C5 DoxA, a cytochrome P-450 catalyzing multiple steps in doxorubicin biosynthesis.* J Bacteriol, 1999. **181**(1): p. 298-304.
121. Hutchinson, C.R., *Biosynthetic Studies of Daunorubicin and Tetracenomycin C.* Chem Rev, 1997. **97**(7): p. 2525-2536.
122. Dickens, M.L., N.D. Priestley, and W.R. Strohl, *In vivo and in vitro bioconversion of epsilon-rhodomycinone glycoside to doxorubicin: functions of DauP, DauK, and DoxA.* J Bacteriol, 1997. **179**(8): p. 2641-50.
123. Bao, W., P.J. Sheldon, E. Wendt-Pienkowski, and C.R. Hutchinson, *The Streptomyces peucetius dpsC gene determines the choice of starter unit in biosynthesis of the daunorubicin polyketide.* J Bacteriol, 1999. **181**(15): p. 4690-5.
124. Bao, W., P.J. Sheldon, and C.R. Hutchinson, *Purification and properties of the Streptomyces peucetius DpsC beta-ketoacyl:acyl carrier protein synthase III that specifies the propionate-starter unit for type II polyketide biosynthesis.* Biochemistry, 1999. **38**(30): p. 9752-7.

125. Ye, J., M.L. Dickens, R. Plater, Y. Li, J. Lawrence, and W.R. Strohl, *Isolation and sequence analysis of polyketide synthase genes from the daunomycin-producing Streptomyces sp. strain C5.* J Bacteriol, 1994. **176**(20): p. 6270-80.
126. Meurer, G., M. Gerlitz, E. Wendt-Pienkowski, L.C. Vining, J. Rohr, and C.R. Hutchinson, *Iterative type II polyketide synthases, cyclases and ketoreductases exhibit context-dependent behavior in the biosynthesis of linear and angular decapolyketides.* Chem Biol, 1997. **4**(6): p. 433-43.
127. Ferrer, J.L., J.M. Jez, M.E. Bowman, R.A. Dixon, and J.P. Noel, *Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis.* Nat Struct Biol, 1999. **6**(8): p. 775-84.
128. Schroder, J., S. Raiber, T. Berger, A. Schmidt, J. Schmidt, A.M. Soares-Sello, E. Bardshiri, D. Strack, T.J. Simpson, M. Veit, and G. Schroder, *Plant polyketide synthases: a chalcone synthase-type enzyme which performs a condensation reaction with methylmalonyl-CoA in the biosynthesis of C-methylated chalcones.* Biochemistry, 1998. **37**(23): p. 8417-25.
129. Suh, D.Y., K. Fukuma, J. Kagami, Y. Yamazaki, M. Shibuya, Y. Ebizuka, and U. Sankawa, *Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases.* Biochem J, 2000. **350 Pt 1**: p. 229-35.
130. Suh, D.Y., J. Kagami, K. Fukuma, and U. Sankawa, *Evidence for catalytic cysteine-histidine dyad in chalcone synthase.* Biochem Biophys Res Commun, 2000. **275**(3): p. 725-30.
131. Suh, D.Y., J. Kagami, K. Fukuma, N. Iwanami, Y. Yamazaki, H. Yurimoto, Y. Sakai, N. Kato, M. Shibuya, Y. Ebizuka, and U. Sankawa, *Chalcone and stilbene synthases expressed in eucaryotes exhibit reduced cross-reactivity in vitro.* Chem Pharm Bull (Tokyo), 2000. **48**(7): p. 1051-4.
132. teaching.shu.ac.uk.
133. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.
134. Gasteiger E., H.C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A, *Protein Identification and Analysis Tools on the ExPASy Server.* John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). , 2005: p. 571-607.
135. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Macromolecular Crystallography, Pt A, 1997. **276**: p. 307-326.
136. Storoni, L.C., A.J. McCoy, and R.J. Read, *Likelihood-enhanced fast rotation functions.* Acta Crystallographica Section D-Biological Crystallography, 2004. **60**: p. 432-438.
137. McCoy, A.J., R.W. Grosse-Kunstleve, L.C. Storoni, and R.J. Read, *Likelihood-enhanced fast translation functions.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 4): p. 458-64.
138. Agarwal, R.C., *A new least-squares refinement technique based on the fast Fourier transform algorithm.* Acta Cryst, 1978. **A34**: p. 791-809.
139. Read, R.J., *Improved Fourier coefficients for maps using phases from partial structures with errors.* Acta Cryst, 1986. **A42**: p. 140-149.

140. Perrakis, A., M. Harkiolaki, K.S. Wilson, and V.S. Lamzin, *ARP/wARP and molecular replacement.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 10): p. 1445-50.
141. Cowtan, K., *`dm`: An automated procedure for phase improvement by density modification* Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, 1994. **31**: p. 34-38.
142. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
143. *The CCP4 suite: programs for protein crystallography.* Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 5): p. 760-3.
144. Laskowski, R.A., M. W. MacArthur, D. S. Moss and J. M. Thornton, *PROCHECK: a program to check the stereochemical quality of protein structures.* J. Appl. Cryst, 1993. **26**(283-291).
145. Davis, I.W., A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, L.W. Murray, W.B. Arendall, 3rd, J. Snoeyink, J.S. Richardson, and D.C. Richardson, *MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W375-83.
146. Davis, I.W., L.W. Murray, J.S. Richardson, and D.C. Richardson, *MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W615-9.
147. Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman, *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.
148. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic Acids Res, 1994. **22**(22): p. 4673-80.
149. www.expasy.org.